

STIC-Biotech/ChemLib

157326

From: Swope, Sheridan  
Sent: Thursday, June 23, 2005 10:08 AM  
To: STIC-Biotech/ChemLib  
Subject: RE: 10/617,433

10/617,443

10617443

Apologies!!

-----Original Message-----

From: STIC-Biotech/ChemLib  
Sent: Thursday, June 23, 2005 6:33 AM  
To: Swope, Sheridan  
Subject: RE: 10/617,433

Good morning, do you have the correct number. 10/617433 is NO BIOTECH DATA.

-----Original Message-----

From: Swope, Sheridan  
Sent: Wednesday, June 22, 2005 5:25 PM  
To: STIC-Biotech/ChemLib  
Subject: 10/617,433

For 10/617,433, pls search:

SID 1: residues 1-1038 oligo search ( $\geq 12$  nts) against the NT databases.

SID 1: residues 1-1038 regular search against the NT databases.

SID 2: residues 1-9 against the NT databases.

SID 2: regular search against the NT databases.

Thanks!!!

Sheridan Swope, Ph.D.  
Patent Examiner, AU 1652  
Recombinant Enzymes  
571-272-0943 (voice)  
E02B71 Remsen Bld (Office)  
E02C70 Remsen Bld (Mailbox)

Barb O'Brien

\*\*\*\*\*

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2-\_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA#: \_\_\_\_\_ AA#: \_\_\_\_\_  
Interference: \_\_\_\_\_ SPDI: \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure#: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: June 29, 2005, 13:51:28 ; Search time 4887 Seconds

(without alignments)  
10291.900 Million cell updates/sec

Title: US-10-617-443B-1\_COPY\_1\_1038

Perfect score: 1038  
Sequence: 1 caggagactcgaagttgcag.....ggccctccgcagcgcag 1038

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:  
1: gb\_da:  
2: gb\_beg:  
3: gb\_in:  
4: gb\_com:  
5: gb\_ov:  
6: gb\_pat:  
7: gb\_ph:  
8: gb\_pl:  
9: gb\_pr:  
10: gb\_ro:  
11: gb\_sta:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vi:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1031.6	99.4	60597	9 AC113611	AC113611 Homo sapi
2	83	8.0	218856	2 AC122871	AC122871 Mus muscu
3	77.2	7.4	235102	2 AC118993	AC118993 Rattus no
4	77.2	7.4	258368	2 AC110369	AC110369 Rattus no
5	77.2	7.4	264075	2 AC134075	AC134075 Rattus no
6	56.4	5.4	125020	9 AF429315	AF429315 Homo sapi
7	56.2	5.4	172105	2 AC021197	AC021197 Homo sapi
8	55.8	5.4	98501	9 AP005019	AP005019 Homo sapi
9	55.8	5.4	111282	6 AR368084	AR368084 Sequence
10	55.8	5.4	111282	6 AR368084	AR368084 Sequence
11	55.8	5.4	194719	2 AC026173	AC026173 Homo sapi
12	55.8	5.4	197022	2 AC055829	AC055829 Homo sapi
13	55.6	5.4	125020	9 AF429315	AF429315 Homo sapi
14	54.8	5.3	173791	9 AC021035	AC021035 Homo sapi
15	53.4	5.1	196512	9 AC020911	AC020911 Homo sapi
16	53.2	5.1	192138	2 AC148545	AC148545 Otolenur
17	52.8	5.1	147454	2 AC027348	AC027348 Homo sapi
18	52.4	5.0	126380	2 AP000714	AP000714 Homo sapi
19	52.4	5.0	159840	2 AP000869	AP000869 Homo sapi

C	20	52.4	5.0	171902	2	AP000846	AP000846 Homo sapi
	21	52.4	5.0	197215	9	AP002954	AP002954 Homo sapi
	22	52.4	5.0	199321	9	AP000941	AP000941 Homo sapi
	23	52.2	5.0	90801	9	AC025033	AC025033 Homo sapi
	24	52.2	5.0	171615	2	AC089988	AC089988 Homo sapi
	25	52.2	5.0	212382	2	AC010936	AC010936 Homo sapi
	26	52	5.0	106216	9	AL161908	AL161908 Human DNA
	27	51	4.9	14725	9	HS1118GA	HS1118GA Human DNA
	28	51	4.9	112560	9	HSJ513G18	HSJ513G18 Human DNA
	29	51	4.9	125111	2	AC147938	AC147938 Oritheorh
	30	51	4.9	212406	2	AC132705	AC132705 Rattus no
	31	51	4.9	217418	2	AC106071	AC106071 Rattus no
	32	51	4.9	244546	2	AC107358	AC107358 Rattus no
	33	50.8	4.9	102591	9	AC002978	AC002978 Homo sapi
	34	50.6	4.9	53269	2	AC006184	AC006184 Homo sapi
	35	50.6	4.9	75031	9	AL390716	AL390716 Human DNA
	36	50	4.8	63266	9	AL607038	AL607038 Human DNA
	37	50	4.8	183672	9	CNS05TC8	CNS05TC8 Human chr
	38	50	4.8	197369	9	CNS01RH6	CNS01RH6 Human chr
	39	49.6	4.8	118495	9	AL590556	AL590556 Human DNA
	40	49.6	4.8	122351	9	AP003780	AP003780 Homo sapi
	41	49.6	4.8	162805	2	AC020565	AC020565 Homo sapi
	42	49.6	4.8	192453	2	AC023920	AC023920 Homo sapi
	43	49.4	4.8	2000	6	AX655393	AX655393 Sequence
	44	49.4	4.8	181309	2	AC150795	AC150795 Bos tauru
	45	49.2	4.7	112911	9	AL355310	AL355310 Human DNA

## ALIGNMENTS

RESULT 1	AC113611	60597 bp	DNA	linear	PRI 10-MAR-2003
LOCUS	AC113611	Homo sapiens BAC clone RP11-421M20	From 4,	complete sequence.	
DEFINITION	AC113611	AC113611.3	GI:28475761	HTG.	
ACCESSION	AC113611				
VERSION	AC113611.3				
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS			</		

## JOURNAL

## COMMENT

Submitted (10-MAR-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Feb 23, 2003 this sequence version replaced gi:20135967.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)

----- Summary Statistics

Center project name: H\_NH0421M20

-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is not the entire insert of the clone. This clone is overlapped by AC104650 and AC104825.

Discrepant bases between AC104650, AC104825 and clone sequence.

Data from AC104825 was used to finish this clone.

## FEATURES

## source

Location/Qualifiers

1..60597

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-421M20"

/clone\_id="RPC1-11"

539..903

/rpt\_family="ERV1"

2797..2907

/rpt\_family="(CAGAGA)n"

2908..3114

/rpt\_family="ERV1"

3122..3184

/rpt\_family="(CAGAGA)n"

3256..3368

/rpt\_family="L1"

3500..3682

/rpt\_family="L1"

3713..3782

/rpt\_family="AT-rich"

5208..5410

/rpt\_family="MIR"

repeat\_region

5451..5503

/rpt\_family="MIR"

repeat\_region

5504..5677

/rpt\_family="MIR"

repeat\_region

6645..6687

/rpt\_family="MIR"

repeat\_region

6998..7067

/rpt\_family="MIR"

repeat\_region

7741..8173

/rpt\_family="MER1\_type"

repeat\_region

8530..8610

/rpt\_family="MIR"

misc\_feature

8638..9514

/note="CpG island (%GC=76.7, o/e=0.80, #CpGs=114)"

repeat\_region

9027..9048

/rpt\_family="GC-rich"

repeat\_region

9063..9111

/rpt\_family="GC-rich"

repeat\_region

9612..9739

/rpt\_family="MIR"

repeat\_region

10416..10509

/rpt\_family="MER103"

repeat\_region

10688..11034

/rpt\_family="MALR"

repeat\_region

11781..11997

/rpt\_family="L2"

repeat\_region

12212..12325

/rpt\_family="MIR"

repeat\_region

13275..13625

/rpt\_family="L2"

repeat\_region

13703..13769

/rpt\_family="CR1"

repeat\_region

14926..15054

/rpt\_family="MIR"

repeat\_region

15239..15281

/rpt\_family="Tlp100"

repeat\_region

15349..15515

/rpt\_family="MIR"

repeat\_region

16000..16113

/rpt\_family="MIR"

repeat\_region

16639..16917

/rpt\_family="L2"

repeat\_region

17127..17228

/rpt\_family="Tlp100"

repeat\_region

17883..17982

/rpt\_family="MIR"

repeat\_region

18108..18173

/rpt\_family="CT-rich"

repeat\_region

19059..19127

/rpt\_family="L2"

repeat\_region

20087..20234

/rpt\_family="MIR"

repeat\_region

20576..20738

/rpt\_family="MIR"

repeat\_region

22242..22380

/rpt\_family="MIR"

repeat\_region

23733..23804

/rpt\_family="(TGAA)n"

repeat\_region

23890..23987

/rpt\_family="L2"

repeat\_region

24011..24440

/rpt\_family="MIR"

repeat\_region

25254..25412

/rpt\_family="MALR"

repeat\_region

26001..26150

/rpt\_family="MIR"

repeat\_region

26473..26512

/rpt\_family="MIR"

repeat\_region

26513..26894

/rpt\_family="MIR"

repeat\_region

26895..27048

/rpt\_family="MALR"

repeat\_region

27423..27724

/rpt\_family="MIR"

repeat\_region





\* runs of N, but the exact sizes of the gaps are unknown.  
\* this record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1554: contig of 1554 bp in length  
\* 1555 1654: gap of unknown length  
\* 1655 2784: contig of 1130 bp in length  
\* 2785 2885: gap of unknown length  
\* 2885 4046: contig of 1162 bp in length  
\* 4047 4146: gap of unknown length  
\* 4147 121446: contig of 117300 bp in length  
\* 121447 121547: gap of unknown length  
\* 121547 218856: contig of 97310 bp in length.

FEATURES  
source  
1.218856  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="UNK"  
/clone="RP23-107P12"  
1.1554  
/note="assembly\_name:Contig9"  
1655.2784  
/note="assembly\_name:Contig15"  
2885.4046  
/note="assembly\_name:Contig17"  
4147.121446  
/note="assembly\_name:Contig20"  
121547.218856  
/note="assembly\_name:Contig21"

## ORIGIN

Query Match 8.0%; Score 83; DB 2; Length 218856;  
Best Local Similarity 60.1%; Pred.No.1.7e-07;  
Matches 295; Conservative 0; Mismatches 160; Indels 36; Gaps 8;  
QY 53 GAGGCGATTTTCAGTCCATTTTTCAGATGAG-GAGTTAGAGCCCGACAGACGTAAAGTAA 111  
DB 200999 GGGGTCACCTTACATGCCATTTGTTAGTGGAGAAATCAAGGCCAGAGAGGTTAACTCA 201058  
QY 112 TCTGTCGAGGCGACACAGACTAGAAAGACAGCCAGGCCAGCCGACCCCTGGTGTGTGCA 171  
DB 201059 CTTGTTACAGGTCACACAGACTAAAGACAGAGGTCAGTCAAGG--CTTAGTTGTGTGCA 201116  
QY 172 GCGCCCGACCCGAGTTCTATTGCGGGGCTCGGAGACCCAGACCGAGCTGAGCAGCATG 231  
DB 201117 GCTCCAGGCGCTG-TGCTCATTTTACATCTCAGGAACTCTCATTAAGCACTGCGCAGGCG 201175  
QY 232 TGTTCAGATGTGGGAACTGGAGAGAGCCCGGACAGGCCCGGTGCGAGGGAACCCCGAGG 291  
DB 201176 TGTG-----GTGACCGAGACCGGGAACCTGAGAGTCTC----- 201207  
QY 292 GCTTAGAGCCCGGTGCGACATGATGCTGAGGCTGTGTGCTCTGCGACCAAGCCCTTA 351  
DB 201208 AGTTTAGGCTCATGTCATCCATGCTGCGGGCTGTGTGCTCTGCGAAGCCACAGCCCTG 201267  
QY 352 CTGCTGACCGGACGACGAGATTTGAGCCCGGAGAG--GCTCAAGGAGAGTTCTGTAACCATG 410  
DB 201268 TTGCTGATGTGGCGCAAAATCTGAGCAGTGAAGAGCTTTGGGGGCTTCGTGTGTGCC 201327  
QY 411 TAGCAAGTGGGGCTGGGGTGTGGCCAAAGTTAGACAGATGTAGGGCCCTGTGAGTCTAG 470  
DB 201328 TGGCAAGTAGGGCTGGGCTATGAGCCAAAGTTGACAAAG--GGGGTTTCTGTCTCACTCAG 201385  
QY 471 AAATTGGCAGCTCTTTTGGCCCAAGAGGGGCAAGCTGTGTTCGGGGCTGGGTAAGTCTAGA 530  
DB 201386 GAATCATTAACCCAGCCAAATAGAGAGGGCTTCACAGAGAGCC--AGCTGAGATGAGCTCAGG 201444  
QY 531 AGGTCACCTG 541  
DB 201445 AGGGGGCCCGAG 201455

## RESULT 3

AC118993/c  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-237H8, WORKING DRAFT SEQUENCE, 2  
unordered pieces.  
AC118993  
AC118993.4 GI:25013258  
VERSION  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTROP.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

## REFERENCE

## AUTHORS

1 (bases 1 to 235102)  
Muzny,D,Marie, Metzker,M,lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alldbrooke,S, Amin,A, Angiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biwako,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Cearar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durkin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C, Falle,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Gante,R, Garcia,A, Garner,T, Garza,M, Gebregorgis,E, Geer,K, Gill,R, Garcia,M, Guerra,W, Guevara,M, Gunaretna,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huliy,S, Hume,J, Idlebird,D, Jackson,A, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenzenhewa,L, Loulseged,H, Lorzado,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindartne,M, Mamoud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mawliny,S, McLeod,M,P, McNeill,T,Z, Meenen,B, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwokedemeh,O, Okwum,G, Olarinmuaogun,A, Pal,S, Parke,K, Paeternak,S, Paul,H, Perez,A, Perez,J, Pfannkoch,C, Plopper,F, Poindexter,A, Popovic,D, Primus,B, Pu,L, L, Puzo,M, Quiroz,J, Rachlin,B, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojase,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savery,G, Scherer,S, Scott,G, Shatman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Stiller,C,D, Smajd,D, Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J, Steimle,M, Strong,R, Sutton,A, Syatel,A, Taboc,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Umami,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,Q, Wang,S, Warren,J, Warren,K, Wel,X, White,F, Williams,G, Willson,R, Wleczyk,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,X, Zhou,X, Zhao,S, Dunn,D, von Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 235102)  
Worley,K.C.  
Direct Submission  
Submitted (24-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 235102)  
Rat Genome Sequencing Consortium.

## REFERENCE

## AUTHORS

TITLE Direct Submission  
JOURNAL Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 15, 2002 this sequence version replaced gi:23581105. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GTNH  
Center clone name: CH230-237H8  
Assembly program: Phrap; version 0.990329

----- Summary Statistics

Consensus quality: 227037 bases at least Q40  
Consensus quality: 229537 bases at least Q30  
Consensus quality: 231225 bases at least Q20  
Estimated insert size: 235071; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbankdraft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 108068: contig of 108068 bp in length  
\* 108069 108168: gap of unknown length  
\* 108169 235102: contig of 126934 bp in length.

FEATURES

source

1. 235102  
/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-237H8"

1. 1143

/note="wgs end extension"

clone end: 5p5'

complement(3003..3706)

/note="clone boundary"

clone end: 5p6'

site:

end sequence: BZ101611"

108169..110729

/note="wgs contig"

ORIGIN

Query Match

Best Local Similarity 58.0%; Pred. No. 2.6e-06;  
Matches 232; Conservative 0; Mismatches 133; Indels 35; Gaps 4;

53 GAGGCGATTTACCTCCATTTTTCAGATGAG-GAGTTGAGGCCCGACAGAGACGTAACTAA 111

89387 GGGGCTCTCTTCCAGTCCACTGTGATGGAAGAGCAAGGCCCGACAGAGGCTAAACGA 89328

112 TCTGTGAGGCCACACAGCTAGAAAGACGACGCCAGCCGCAACCCCTGTGTGTGCA 171

Db

89327 CTTGTTCAGGGCAGACAGCTAAAGAAAGACAGAGCTCCAGCTGGA--CCAGTTGTGTGCA 89270

Qy

172 GCGCCCGACCCAGTTGTCTCAATTCGCGGGGCTCGGGAGCCAGAGGAGCTGAGAGCATG 231

Db

89269 GCTCTCAGTCTCTGATCTATTTAGACTTCAGGAGACCTCTCATAGACCTGCCAAGCT 89210

Qy

232 TGTTCAGATGTGGGGAAGTGAAGAGAGCCCGGACAGAGCCCGTGACAGGAAACCCGAG 291

Db

89209 GTGTGCGCAGAGACCGGAGCCTGAGAGCTTCAGTACAGGCTCATGT----- 89164

Qy

292 GCTGTAGGCCCCGTCGACATGCTCATGCGCTGTGTGTCTTGTGACCAAGCCCTTA 351

Db

89163 -----CACCCAGATGCTCAGAGCCTGTGTGTCTTGTGACCAAGCCCTG 89119

Qy

352 CTGTGACGCGACAGAGAAATCTGAGCCCG--GGAAGGTCGAGGAAATTCGTAACCA 408

Db

89118 TTGCTGATGTGTGGAGAAATCTGAGCCAGATGATGAAGTTGGGAGAGTTTCATGTCTG 89059

Qy

409 TCTAGCAAGTCGGGCTGGGGTGGCCAGCTTAGACACAG 448

Db

89058 CTTAGCAAGTAGGCTTGGGATATGGCCAGATTGACAAAG 89019

RESULT 4

AC110369/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 258368)

Muzny,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Angiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Cessari,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huylk,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,D, Lorenshuwa,L, Loullesed,H, Lozada,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapa,P, Martin,K, Martin,R, Martinez,E, Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Munidas,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Naackelemehe,O, Okunonu,G, Olarnunagsoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Plankoch,C, Plopper,F, Polindexter,A, Popovic,D, Primus,E, Pu,L, Pu,M, Puazo,M, Quiroz,D, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,

AC110369

AC110369

HTG: HTG PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

AC110369

AC110369

HTG: HTG PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

AC110369

AC110369

HTG: HTG PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

AC110369

AC110369

HTG: HTG PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

AC110369

AC110369

HTG: HTG PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

**TITLE**  
JOURNAL  
REFERENCE  
**AUTHORS**  
**TITLE**  
JOURNAL  
REFERENCE  
**AUTHORS**  
**TITLE**  
JOURNAL  
COMMENT

Relly, B., Rodkey, T., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rivers, C., Rojars, A., Rose, M., Rose, R., Ruiz, S.J.,  
Sanders, W., Savery, G., Scheer, S., Scott, G., Shatman, S., Shen, H.,  
Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Snajd, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorille, R., Sosa, J.,  
Steinle, M., Strong, R., Suton, A., Svatek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tinney, A., Trejos, Z., Usami, K.,  
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., Wiley, F.,  
Williams, G., Willison, R., Wleczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von  
Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.  
Direct Submission  
2 (bases 1 to 258368)  
Unpublished  
Mortley, K.C.  
Submitted (11-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 258368)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Sep 19, 2002 this sequence version replaced g1:21744398.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the  
sequence may extend beyond the ends of the clone and there may be  
contigs that consist entirely of whole genome shotgun sequence  
reads. Both end sequences and whole genome shotgun sequence only  
contigs will be indicated in the feature table.

--Genome Center--  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
-----Project Information-----  
Center project name: GRP  
Center clone name: CH230-49K10  
-----Summary Statistics-----  
Assembly program: Phrap, version 0.990329  
Consensus quality: 236348 bases at least Q40  
Consensus quality: 238877 bases at least Q30  
Consensus quality: 240713 bases at least Q20  
Estimated insert size: 259061; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation  
-----  
NOTE: Estimated insert size may differ from sequence length  
[see http://www.hgsc.bcm.tmc.edu/docs/genbank\_drift\_data.html]  
NOTE: This sequence may represent more than one clone.  
NOTE: This is a 'working draft' sequence. It currently  
consists of 3 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 102159: contig of 102159 bp in length  
\* 102160 102259: gap of unknown length  
\* 102260 249257: contig of 146998 bp in length  
\* 249358 249357: gap of unknown length  
\* 249358 258368: contig of 9011 bp in length.  
Location/Qualifiers  
1 . 258368  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"

**FEATURES**  
source

[illegible]



ACCESSION AF429315  
VERSION AF429315.1 GI:17646244  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
REFERENCE 1 (bases 1 to 125020)  
AUTHORS Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S.,  
Ingersoll-Ashworth, R.G., Fleisher, A., Stevalin, G., Brice, A.,  
Potter, N.T., Ross, C.A. and Margolis, R.L.  
TITLE A repeat expansion in the gene encoding juncctophilin-3 is  
associated with Huntington disease-like 2  
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)  
MEDLINE 21583737  
PUBMED 11694876  
REFERENCE 2 (bases 1 to 125020)  
AUTHORS Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Psychiatric, Johns Hopkins Medical  
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA  
FEATURES  
source  
1. 125020  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16q24.3" between D16S520 and WI-12410"  
/note="Isolated from a patient with Huntington's  
disease-like 2 (HDL2)"  
complement(135581..35746)  
/rpt\_type="tandem"  
/rpt\_unit="ctg"  
complement(<36507..>36887)  
/gene="JPH3"  
/note="synonym: JPH3"  
complement(<36507..>36887)  
/gene="JPH3"  
/product="juncctophilin 3"  
complement(<36507..36887)  
/gene="JPH3"  
/note="component of the junctional complex between plasma  
membrane and endoplasmic reticulum"  
/codon\_start=1  
/product="juncctophilin 3"  
/protein\_id="AA140941.1"  
/db\_xref="GI:17646245"  
/translation="MSSGGRFNFDGSGYCGMEDKAKHGIVCTGPKGGEYTGSHS  
HGFELVGLYTPSGNTGCTGTAQGRHGI.GLESKGKVVYKGMTHGFGRYGVECA  
NGAKYBGTWSNGLQDGYETRYSDG"  
ORIGIN  
Query Match 5.4%; Score 56.4; DB 9; Length 125020;  
Best Local Similarity 10.8%; Pred. No. 0.058; Indels 14; Gaps 4;  
Matches 95; Conservative 398; Mismatches 376;  
127 ACAGCTAGAAAGCAGCAGCCAGCCAGCAACCCCTGCTGTCAGACCCCAAGCTT 186  
16895 MSMMMSVSYVKMSMSASBSCHMBWCKMTWSCCMNMTKSSWMSWGCMGRRSRK 16954  
187 GCTGATTGCGGGCTCGGAGCCAGCAGGAGGCTGAGCACTGTGTTCCAGATGTGG 246  
16955 WKMSRMSRSKSRMYRYSWMSWMTSSCYASWCMCCWSCCMSCCMSCCYSC 17014  
247 GAAGTGGAGAGAGCCCGGACAGCCGCTGAGGAAACCCGAGGGGCTGAGAGCCCGT 306  
17015 CMRYCCACCKCYMSWTWMSASYSRSTYKSMKCMKSRSRSRKSGRGSGMSGG 17074  
307 CCACCTGCATGCTGAGGCTGCTGCTGAGCAGCAGCCCTACTGCTGAGGAGCA 366  
17075 KGGSYGKTKRKSRGKAKKMYVYRSRKMKNYSSKGMVCMYCMGRRGCTYCSM 17134  
367 GGAATCTGAGCCCGGAAAGGTCCAGGGAATT--CGTGAACATCTAGCAAGTGGGCT 424

Db 17135 TSSAMCCSYCAKCKSKCYCYMSKGYVYACYSRGSMSKXCMRSGTSTSCCCC 17194  
425 GGGGTGTGGCCAAAGTTAGACAAGATGAGGCGCCCTGTGACTAGAAATTTGAGCTT 484  
17195 TTTTTCCTCCCNANTGGAGAGCTTTTNCNKTYSYRRRNCACCKYNNYNSWRSR 17254  
485 TTTTGCCCAAGAGGGCCACCGCTGTGCGGGCTGGTAGCTCAGAAAGTCACTTGGG 544  
17255 AGMSTKRYKSWTMSASBSCHMBWCKMTWSCCMNMTKSSWMSWGCMGRRSRK 17314  
545 GTCTTCCACTACA---CCCCGCGCTGACACTCTGTAGCCCCAGGGCTGGAGGAGC 600  
17315 WCKMKCYCMRMSRSKSGMYVYASWKSRRCTCTRCYCMSSSKYKSYMMRSGM 17374  
601 AGCTGAGCCCATGAGAGAGGCGCCAGTTCTCTCCCTGTAAGGATTTCTGTAGCATG 660  
17375 KGMKSRKMSWMSASBSCHMBWCKMTWSCCMNMTKSSWMSWGCMGRRSRK 17434  
661 GGAACAGACAGAGCCAGGAGGACTAACCCGAGATCCAGCCCGGCTTACTCCGCTG 720  
17435 RRMWKSAMTMRMRMWRGAMMMCKKCYSPMSRCMMWKSICYASCRCSCAMMSGTYMKA 17494  
721 GCTCAGCGCAATATCTTAACCTCTCTGTGAGCCTCTGCCAGCTAGAGGATCCAGT 780  
17495 SYMCMSSYWCCTCWTCTSYSTRCTCCKGMSITTKSSMSMSYKGRRSYSWCT 17554  
781 AG-----GGGGGTGAGAGCCAGCAGCAGTGAAGCTTTTAAACATTTCCGGGTA 834  
17555 KGMKSRKMSWMSASBSCHMBWCKMTWSCCMNMTKSSWMSWGCMGRRSRK 17614  
835 GCGAGCCCTTCCAAATGCTGTGTACATGCACTGTGTGTGATGGGGTCCCCAAC 894  
17615 MMTWYCCCMRRRSYVYRSMSAMGRKSSWSGMMMSASRRCKASRSWSRBMKG 17674  
895 GGGCTCAGTGTGGGTGAGGCTGCTGCTGAACCTGG--ACAGGGCTCAGAGAGCTT 952  
17675 RSCWSSKMGWSKRSASBSCHMBWCKMTWSCCMNMTKSSWMSWGCMGRRSRK 17734  
953 CCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 995  
17735 CWRKSGSMTSSCYYSASSCWMMMSKSCMCCMMKRRCAAC 17777  
RESULT 7  
AC021197 172105 bp DNA linear HTG 07-JUN-2000  
LOCUS Homo sapiens chromosome 19 clone RP11-798B19, WORKING DRAFT  
DEFINITION  
AC021197  
AC021197.5 GI:8570404  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 172105)  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Waterston, R.H.  
AUTHORS Unpublished  
JOURNAL 2 (bases 1 to 172105)  
REFERENCE 2 (bases 1 to 172105)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT  
On Jun 17, 2000 this sequence version replaced gi:7235351.  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----



```
Center project name: H.NH0798B19
----- Summary Statistics -----
Sequencing vector: MJ3; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160434 bases at least Q40
Consensus quality: 164353 bases at least Q30
Consensus quality: 166763 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 170105; sum-of-contigs
Quality coverage: 3.66 in Q20 bases; agarose-fp
Quality coverage: 4.02 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
1301: contig of 1301 bp in length
1302
1401: gap of unknown length
1402
4060: contig of 2659 bp in length
4061
4060: gap of unknown length
4062
7024: contig of 2864 bp in length
7025
7124: gap of unknown length
7125
9662: contig of 2538 bp in length
9663
9762: gap of unknown length
9763
13070: contig of 3308 bp in length
13071
13170: gap of unknown length
13171
15928: contig of 2758 bp in length
15929
16028: gap of unknown length
16029
19294: contig of 3266 bp in length
19295
19394: gap of unknown length
19395
23283: contig of 3889 bp in length
23284
23883: gap of unknown length
23884
27272: contig of 3889 bp in length
27273
27372: gap of unknown length
27373
32535: contig of 5163 bp in length
32536
40106: gap of unknown length
40107
40206: gap of unknown length
40207
45656: contig of 5450 bp in length
45657
52232: contig of 6476 bp in length
52233
52332: gap of unknown length
52333
59204: contig of 6872 bp in length
59205
59304: gap of unknown length
59305
68409: contig of 9105 bp in length
68410
68509: gap of unknown length
68510
76838: contig of 8329 bp in length
76839
76938: gap of unknown length
76939
88902: contig of 11964 bp in length
88903
89002: gap of unknown length
89003
100235: contig of 11233 bp in length
100236
100335: gap of unknown length
100336
117760: contig of 17425 bp in length
117761
117860: gap of unknown length
117861
142316: contig of 24456 bp in length
142317
142416: gap of unknown length
142417
172105: contig of 29689 bp in length.
Location/Qualifiers
1. 172105
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-798B19"
misc_feature
1. 1301
/note="assembly_name:Contig11"
```

```
misc_feature 1402..4060
/note="assembly_name:Contig12"
misc_feature 4161..7024
/note="assembly_name:Contig13"
misc_feature 7125..9662
/note="assembly_name:Contig14"
misc_feature 9763..13070
/note="assembly_name:Contig15"
misc_feature 13171..15928
/note="assembly_name:Contig16"
misc_feature 16029..19294
/note="assembly_name:Contig17"
misc_feature 19395..23283
/note="assembly_name:Contig18"
misc_feature 23384..27272
/note="assembly_name:Contig19
clone_end:SP6
vector_side:right"
misc_feature 27373..32535
/note="assembly_name:Contig20"
misc_feature 32636..40106
/note="assembly_name:Contig21"
misc_feature 40207..45656
/note="assembly_name:Contig22"
misc_feature 45757..52332
/note="assembly_name:Contig23"
misc_feature 52333..59204
/note="assembly_name:Contig24"
misc_feature 59305..68409
/note="assembly_name:Contig25"
misc_feature 68510..76838
/note="assembly_name:Contig26"
misc_feature 76939..88902
/note="assembly_name:Contig27"
misc_feature 89003..100235
/note="assembly_name:Contig28"
misc_feature 100336..117760
/note="assembly_name:Contig29"
misc_feature 117861..142316
/note="assembly_name:Contig30"
misc_feature 142417..172105
/note="assembly_name:Contig31"

ORIGIN
Query Match 5.4%; Score 56.2; DB 2; Length 172105;
Best Local Similarity 65.6%; Pred. No. 0.058;
Matches 82; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
QY 55 GGGCATTTGAGCCCAATTTTTCAGATGAGAGTGGCCGAGAGAACTAATCT 114
DB 103864 GGCTGATGTTATTTCCATTTTACAGAGAGGGGTTGAGGCCGAGAGAACTT 103923
QY 115 GTCTGAGCCACACAGCTAGAAAGCAGCCGAGCCGAGCCGAGCCGAGCCGAGCC 174
DB 103924 GCCTGAAGCCACACAGAGCCGAGCCGAGCCGAGCCGAGCCGAGCCGAGCC 103983
QY 175 CCACAG 179
DB 103984 CTCAG 103988

RESULT 8
AP005019 98501 bp DNA linear PRI 30-MAR-2002
LOCUS Homo sapiens genomic DNA, chromosome 11q clone:RP11-169D4, complete
DEFINITION
ACCESSION AP005019
VERSION AP005019.1 GI:19879835
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
```

REFERENCE 1  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Homo sapiens genomic DNA  
JOURNAL Published Only in Database (2002)  
REFERENCE 2 (bases 1 to 98501)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAR-2002) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.sec.riken.go.jp/  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
FEATURES  
source Location/Qualifiers  
1.98501  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/map="11q"  
/clone="RP11-169D4"  
ORIGIN  
Query Match 5.4%; Score 55.8; DB 9; Length 98501;  
Best Local Similarity 70.1%; Pred. No. 0.083;  
Matches 89; Conservative 0; Mismatches 37; Indels 1; Gaps 1;  
QY 67 TCCCATTTTCAGATGAGAGTTGAGGCCCGAGAACTGAATCTGTCTGAGGCCAC 126  
Db 77268 TCCCATTTTCAGAGTGTGAAGCTGAGGCCCGAGAGAGCTTGTGACTGCTCAAGGTCAC 77327  
QY 127 ACAGCTAGAAAGACGACCGCC-CAGCCGAAACCCCTGTGTGTGAGCCCCGACCCCACT 185  
Db 77328 ACAGCTCAAAAGACGAAAGCCGAAAGCTGAACCCAGTCTGTCCAGCTCCCTGGCCAGG 77387  
QY 186 TGCTCAT 192  
Db 77388 CCATCTT 77394  
RESULT 9  
AR368084/c 111282 bp DNA linear PAT 12-SEP-2003  
LOCUS AR368084  
DEFINITION Sequence 3 from patent US 6376225.  
ACCESSION AR368084  
VERSION AR368084.1 GI:34601657  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 111282)  
AUTHORS Wei,M.-H., Wang,X., Merkulov,G.V., Di Francesco,V. and Beasley,E.M.  
TITLE Isolated human phosphodiesterase proteins, nucleic acid molecules  
JOURNAL encoding human phosphodiesterase proteins, and uses thereof  
FEATURES  
source Location/Qualifiers  
1.111282  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 5.4%; Score 55.8; DB 6; Length 111282;  
Best Local Similarity 70.1%; Pred. No. 0.08; Mismatches 37; Indels 1; Gaps 1;  
Matches 89; Conservative 0; Mismatches 37; Indels 1; Gaps 1;  
QY 67 TCCCATTTTCAGATGAGAGTTGAGGCCCGAGAACTGAATCTGTCTGAGGCCAC 126  
Db 93162 TCCCATTTTCAGAGTGTGAAGCTGAGGCCCGAGAGAGCTTGTGACTGCTCAAGGTCAC 93103  
QY 127 ACAGCTAGAAAGACGACCGCC-CAGCCGAAACCCCTGTGTGTGAGCCCCGACCCCACT 185  
Db 93102 ACAGCTCAAAAGACGAAAGCCGAAAGCTGAACCCAGTCTGTCCAGCTCCCTGGCCAGG 93043  
QY 186 TGCTCAT 192  
Db 93042 CCATCTT 93036

QY 186 TGCTCAT 192  
Db 93042 CCATCTT 93036  
RESULT 10  
AX662038/c 111282 bp DNA linear PAT 22-MAR-2003  
LOCUS AX662038  
DEFINITION Sequence 3 from Patent WO02053718.  
ACCESSION AX662038  
VERSION AX662038.1 GI:29163011  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1  
AUTHORS Wei,M.-H., Wang,X., Merkulov,G., di Francesco,V. and Beasley,E.M.  
TITLE Isolated human phosphodiesterase proteins, nucleic acid molecules  
JOURNAL encoding human phosphodiesterase proteins, and uses thereof  
FEATURES  
source Location/Qualifiers  
1.111282  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 5.4%; Score 55.8; DB 6; Length 111282;  
Best Local Similarity 70.1%; Pred. No. 0.08;  
Matches 89; Conservative 0; Mismatches 37; Indels 1; Gaps 1;  
QY 67 TCCCATTTTCAGATGAGAGTTGAGGCCCGAGAACTGAATCTGTCTGAGGCCAC 126  
Db 93162 TCCCATTTTCAGAGTGTGAAGCTGAGGCCCGAGAGAGCTTGTGACTGCTCAAGGTCAC 93103  
QY 127 ACAGCTAGAAAGACGACCGCC-CAGCCGAAACCCCTGTGTGTGAGCCCCGACCCCACT 185  
Db 93102 ACAGCTCAAAAGACGAAAGCCGAAAGCTGAACCCAGTCTGTCCAGCTCCCTGGCCAGG 93043  
QY 186 TGCTCAT 192  
Db 93042 CCATCTT 93036  
RESULT 11  
AC026173 194719 bp DNA linear HTG 05-APR-2002  
LOCUS AC026173  
DEFINITION Homo sapiens chromosome 3 clone RP11-169D4 map 3p, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 62 unordered pieces.  
ACCESSION AC026173  
VERSION AC026173.3 GI:20042926  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 194719)  
AUTHORS Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,  
Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,  
He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,  
Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,  
Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,  
Song,S., Sun,W., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,  
Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,  
Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X.,  
Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,  
Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,  
Yu,J. and Yang,H.  
TITLE Chromosome 3p genomic sequence  
JOURNAL Unpublished



REFERENCE  
AUTHORS  
2 (bases 1 to 194719)  
Chen,J., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B.,  
Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y.,  
Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y.,  
Li,G., Li,C., Bao,Q., Bao,J., Song,L., Zhang,L., Guo,D.,  
Huang,F., Zhang,G., Li,J., Bian,X., Zhang,W., Li,L., Feng,X., Yu,J.  
and Yang,H.  
TITLE  
JOURNAL  
Direct Submission  
Submitted (21-MAR-2000) Human Genomic Center, Institute of  
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
100101, P.R.China  
3 (bases 1 to 194719)  
3 (bases 1 to 194719)  
REFERENCE  
AUTHORS  
Bao,W., Sun,Y., Wu,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,  
Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,  
He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,  
Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,  
Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,  
Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,  
Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,  
Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,  
Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,  
Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,  
Yu,J., and Yang,H.  
TITLE  
JOURNAL  
Direct Submission  
Submitted (05-APR-2002) Human Genomic Center, Institute of  
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
100101, P.R.China  
On Apr 5, 2002 this sequence version replaced gi:8101143.  
-----Genome Center  
Center:Beijing Center  
Center code:Beijing  
Website:http://hgsc.igtp.ac.cn  
http://www.genomics.org.cn  
Contact:hgsc@igtp.ac.cn  
----- Project Information  
Center project name:1% project  
Center clone name: RP11-169D4  
----- Summary Statistics  
Sequencing vector: pUC18, 100% of reads  
Chemistry: Dye-terminator: ET 5% of reads  
Chemistry: Dye-terminator Big Dye; 4% of reads  
Assembly program: Phrap; version 0.99039  
Consensus quality: 37507 bases at least Q40  
Consensus quality: 45201 bases at least Q30  
Consensus quality: 51060 bases at least Q20  
Insert size: 52554, sum-of-contigs  
Quality coverage: 1.65x in Q20 bases;sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 62 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1336: contig of 1336 bp in length  
\* 1337 1436: gap of unknown length  
\* 1437 2122: contig of 686 bp in length  
\* 2123 2223: gap of unknown length  
\* 2223 2945: contig of 723 bp in length  
\* 2946 3045: gap of unknown length  
\* 3046 3860: contig of 815 bp in length  
\* 3861 4756: contig of 796 bp in length  
\* 4757 4856: gap of unknown length  
\* 4857 5429: contig of 573 bp in length  
\* 5430 6114: gap of unknown length  
\* 6114 6214: gap of unknown length  
\* 6215 7064: contig of 850 bp in length  
\* 7065 7164: gap of unknown length  
\* 7165 7982: contig of 818 bp in length  
\* 7983 8082: gap of unknown length  
\* 8083 8795: contig of 713 bp in length  
\* 8796 8895: gap of unknown length  
\* 8896 9588: contig of 693 bp in length  
\* 9589 9688: gap of unknown length  
\* 9689 10512: contig of 824 bp in length  
\* 10513 10612: gap of unknown length  
\* 10613 11424: contig of 811 bp in length  
\* 11424 11523: gap of unknown length  
\* 11524 11524: gap of unknown length  
\* 11524 12509: gap of 985 bp in length  
\* 12509 12608: gap of unknown length  
\* 12608 13724: contig of 1116 bp in length  
\* 13724 13824: gap of unknown length  
\* 13825 14905: contig of 1081 bp in length  
\* 14906 15005: gap of unknown length  
\* 15006 16241: contig of 1235 bp in length  
\* 16241 16341: gap of unknown length  
\* 16341 17953: contig of 1612 bp in length  
\* 17953 18053: gap of unknown length  
\* 18053 19288: contig of 1235 bp in length  
\* 19288 19388: gap of unknown length  
\* 19388 20388: contig of 1001 bp in length  
\* 20388 20488: gap of unknown length  
\* 20489 21703: contig of 1215 bp in length  
\* 21704 21803: gap of unknown length  
\* 21804 22654: contig of 851 bp in length  
\* 22655 22754: gap of unknown length  
\* 22755 24998: contig of 2244 bp in length  
\* 24999 25099: gap of unknown length  
\* 25099 26465: contig of 1367 bp in length  
\* 26466 26565: gap of unknown length  
\* 26566 28332: contig of 1767 bp in length  
\* 28333 28433: gap of unknown length  
\* 28433 29906: contig of 1474 bp in length  
\* 29907 30006: gap of unknown length  
\* 30007 32797: contig of 2791 bp in length  
\* 32798 32897: gap of unknown length  
\* 32898 35022: contig of 2125 bp in length  
\* 35023 35122: gap of unknown length  
\* 35123 37579: contig of 2457 bp in length  
\* 37580 37679: gap of unknown length  
\* 37680 39880: contig of 2201 bp in length  
\* 39881 39980: gap of unknown length  
\* 39981 42263: contig of 2283 bp in length  
\* 42264 42363: gap of unknown length  
\* 42364 44571: contig of 2208 bp in length  
\* 44572 44671: gap of unknown length  
\* 44672 48147: contig of 3475 bp in length  
\* 48147 48246: gap of unknown length  
\* 48247 50501: contig of 2255 bp in length  
\* 50502 50601: gap of unknown length  
\* 50602 52997: contig of 2396 bp in length  
\* 52998 53097: gap of unknown length  
\* 53098 56547: contig of 3450 bp in length  
\* 56548 56647: gap of unknown length  
\* 56648 60181: contig of 3534 bp in length  
\* 60182 60281: gap of unknown length  
\* 60282 63180: contig of 2899 bp in length  
\* 63181 63280: gap of unknown length  
\* 63281 66358: contig of 3078 bp in length  
\* 66359 66458: gap of unknown length  
\* 66459 67956: contig of 1498 bp in length  
\* 67957 68056: gap of unknown length  
\* 68057 71612: contig of 3556 bp in length  
\* 71613 71713: gap of unknown length  
\* 71713 74533: contig of 2820 bp in length  
\* 74533 74633: gap of unknown length  
\* 74633 78476: gap of 3844 bp in length  
\* 78477 78576: gap of unknown length  
\* 78577 82244: contig of 3668 bp in length  
\* 82245 82345: gap of unknown length  
\* 82346 86136: contig of 3791 bp in length  
\* 86137 86235: gap of unknown length

```

*      86236      90642: contig of 4407 bp in length
*      90742      90742: gap of unknown length
*      90743      95083: contig of 4341 bp in length
*      95084      95183: gap of unknown length
*      95184      98380: contig of 3197 bp in length
*      98381      98480: gap of unknown length
*      102415     102414: contig of 3934 bp in length
*      102515     102514: gap of unknown length
*      106775     106775: contig of 4261 bp in length
*      106776     106875: gap of unknown length
*      112291     112291: contig of 5416 bp in length
*      112292     112391: gap of unknown length
*      116740     116740: contig of 4349 bp in length
*      116741     116840: gap of unknown length
*      121741     121741: contig of 4901 bp in length
*      121841     121841: gap of unknown length
*      127122     127122: contig of 5281 bp in length
*      127123     127222: gap of unknown length
*      132291     132291: contig of 5069 bp in length
*      132392     132391: gap of unknown length
*      139100     139100: contig of 6709 bp in length
*      139201     139200: gap of unknown length
*      148094     148094: contig of 8894 bp in length
*      148095     148194: gap of unknown length
*      148195     155530: contig of 7336 bp in length
*      155531     155630: gap of unknown length
*      155631     163081: contig of 7451 bp in length
*      163082     163181: gap of unknown length
*      163182     171479: contig of 8298 bp in length
*      171480     171479: gap of unknown length

```

Query Match Best Local Similarity 70.1%; Score 55.8; DB 2; Length 194719;

Matches 89; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

```

Qy      67      TCCCATTTTCAGTGTGAGGAGTTCAGGCGCCAGAGAACTAATCTGTCTGAGGCCAC 126
Db      44790    TCCCATTTTCAGTGTGAGGAGTTCAGGCGCCAGAGAACTAATCTGTCTGAGGCCAC 44849

```

```

Qy      127     ACAGCTGAAGAGCAGCCAGGCGC-CAGCCGAACCCCTGTGTGTGACAGCCCGCCAGCCAGT 185
Db      44850    ACAGCTGAAGAGCAGCCAGGCGCAGGCTGTGTGTGACAGCCCGCCAGG 44909

```

```

Qy      186     TGCTCAT 192
Db      44910    CCATCTT 44916

```

```

RESULT 12
AC055829 197022 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 11 clone RP11-796A3 map 11, WORKING DRAFT
DEFINITION
AC055829
SEQUENCE 28 unordered pieces.
AC055829 3 GI:9966988
KEYWORDS HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 11, clone RP11-796A3
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 197022)
REFERENCES Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Baetsen, V., Beda, F.,
Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G.,
Bougenbavsky, L., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,
Campopiano, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S.,
Collins, S., Collamore, A., Cooke, P., Dearrellano, K., Dewar, K.,
Dodge, S., Domano, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,

```

# TITLE JOURNAL REFERENCE AUTHORS

## TITLE JOURNAL COMMENT

Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lechoczy, J., Levine, R., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R., Meldrum, D., Menus, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Piant, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teafaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 197022)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Baetsen, V., Boguslavsky, L., Bouckhalter, B., Brown, A., Campopiano, A., Chang, J., Choquel, Y., Colangelo, M., Camarata, J., Campopiano, A., Cooke, P., Dearrellano, K., Dewar, K., Collins, S., Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T., Lechoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Margulis, N., Matthews, C., McCarthy, M., McKernan, K., McPherson, R., Meldrum, D., Menus, L., Mihova, T., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Sougnuez, C., Spencer, R., Subramanian, A., Talamas, J., Teafaye, S., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teafaye, S., Theodore, J., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information  
Center project name: L8652  
Center clone name: 796 A.3  
Summary Statistics  
Sequencing vector: M13, M77815: 96% of reads  
Sequencing vector: Plasmid, n/a; 0.0% of reads  
4.23053341508277Chemistry: Dye-terminator Big Dye, 100% of reads

Assembly program: Phrap, version 0.960721  
Consensus quality: 183643 bases at least Q40  
Consensus quality: 189993 bases at least Q20  
Consensus quality: 192582 bases at least Q20  
Insert size: 193000; agarose-fp  
Insert size: 194322; sum-of-contigs  
Quality coverage: 4.5 in Q20 bases; agarose-fp  
Quality cov.

NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will

```
* be preserved.
1
11811 11810: contig of 11810 bp in length
11911 11910: gap of 100 bp
13354 13354: contig of 1444 bp in length
13355 13354: gap of 100 bp
13455 13454: gap of 100 bp
15818 15818: contig of 2364 bp in length
15819 15818: gap of 100 bp
15919 15918: gap of 100 bp
18437 18437: contig of 2519 bp in length
18438 18437: gap of 100 bp
18537 18537: gap of 100 bp
18538 18537: gap of 100 bp
20676 20675: contig of 2138 bp in length
20776 20775: gap of 100 bp
22055 22054: contig of 1279 bp in length
22155 22154: gap of 100 bp
25416 25415: contig of 3261 bp in length
25516 25515: gap of 100 bp
27880 27880: contig of 2265 bp in length
27881 27880: gap of 100 bp
27981 27980: gap of 100 bp
30860 30859: contig of 2879 bp in length
30960 30959: gap of 100 bp
35487 35486: contig of 4527 bp in length
35587 35586: gap of 100 bp
38315 38314: contig of 2728 bp in length
38415 38414: gap of 100 bp
42075 42074: contig of 3660 bp in length
42175 42174: gap of 100 bp
45912 45912: contig of 3738 bp in length
46013 46012: gap of 100 bp
46013 46012: gap of 100 bp
51705 51704: gap of 100 bp
51705 51704: gap of 100 bp
57385 57385: contig of 5681 bp in length
57486 57485: gap of 100 bp
63567 63567: contig of 6082 bp in length
63568 63567: gap of 100 bp
70096 70096: contig of 6429 bp in length
70197 70196: gap of 100 bp
77591 77590: contig of 7394 bp in length
77691 77690: gap of 100 bp
82359 82359: contig of 4569 bp in length
82360 82359: gap of 100 bp
89812 89812: contig of 7453 bp in length
89913 89912: gap of 100 bp
99111 99110: contig of 9198 bp in length
99211 99210: gap of 100 bp
108487 108486: contig of 9276 bp in length
108587 108586: gap of 100 bp
115026 115025: contig of 6439 bp in length
115126 115125: gap of 100 bp
140162 140161: contig of 25036 bp in length
140162 140161: gap of 100 bp
140262 140261: gap of 100 bp
150678 150677: contig of 10416 bp in length
150778 150777: gap of 100 bp
163255 163255: contig of 12478 bp in length
163256 163255: gap of 100 bp
166098 166098: contig of 22743 bp in length
186099 186198: gap of 100 bp
186199 186198: gap of 100 bp
186199 197022: contig of 10824 bp in length.

FEATURES
Source
1. 197022
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-796A3"
/clone_id="RP11 Human Male BAC"
1. 11810
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
11911. 13354
/note="assembly_fragment"
13455. 15818
/note="assembly_fragment"

misc_feature
15919. 18437
/note="assembly_fragment"
18538. 20675
/note="assembly_fragment"
20776. 22054
/note="assembly_fragment"
22155. 25415
/note="assembly_fragment"
25516. 27880
/note="assembly_fragment"
27981. 30859
/note="assembly_fragment"
30960. 35486
/note="assembly_fragment"
35587. 38314
/note="assembly_fragment"
38415. 42074
/note="assembly_fragment"
42175. 45912
/note="assembly_fragment"
46013. 51604
/note="assembly_fragment"
51705. 57385
/note="assembly_fragment"
57486. 63567
/note="assembly_fragment"
63668. 70096
/note="assembly_fragment"
70197. 77590
/note="assembly_fragment"
77691. 82259
/note="assembly_fragment"

Query Match 5.4%; Score 55.8; DB 2; Length 197022;
Best Local Similarity 70.1%; Pred. No. 0.067;
Matches 89; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

Qy 67 TCCATTTCAGATGAGGTGAGGCCAGAGACCTAGTATCTGTGTGAGGCGAC 126
Db 160668 TCCCATTTTCAGGTGAGTAAGTCCAGAGAGGTTAGTACTTCTCAAGGTAC 160627

Qy 127 ACAGCTGAAGAAGCCAGCCAGCC-C-CAGCCGAACCCCTGTGTGTCAGCCCCAGCCAGT 185
Db 160628 ACAGCTGAAGAAGCCAGCCAGCCAGCCAGCTGTGTCTCCAGCTCCCTGCCAGG 160687

Qy 186 TGCTCAT 192
Db 160688 CCATCTT 160694

RESULT 13
AF429315/c 125020 bp DNA linear PRI 18-JAN-2002
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 125020)
Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
AUTHORS Direct Submission
TITLE Submitted (05-OCT--2001) Psychiatry, Johns Hopkins Medical
```

FEATURES  
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA  
Location/Qualifiers

source

1..125020 "Homo sapiens"  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16q24.3; between D16S20 and WI-12410"  
/note="isolated from a patient with Huntington's  
disease-like 2 (HDL2)"  
complement(13581..35746)  
/rpt\_type="cdg"  
/rpt\_unit="ctg"  
complement(<36507..>36887)  
/gene="JPH3"  
/note="synonym: JP3"  
complement(<36507..>36887)  
/gene="JPH3"  
/product="junctional protein 3"  
complement(<36507..36887)  
/note="component of the junctional complex between plasma  
membrane and endoplasmic reticulum"  
/codon\_start=1  
/product="junctional protein 3"  
/protein\_id="P41409.1"  
/db\_xref="GI:17664245"  
/translation="MSGGRRPNPDGGSYGGGMDKAGHGVCTGPKGQEGYTSMS  
HGFEVLGYTPSGNTYQGTWAGKRGIGLBSKGVYNGEYTHFVKRYGVRECA  
NGAKYEGTWSNGLDGYETETSDG"

mRNA

CDS

5.4%; Score 55.6; DB 9; Length 125020;  
Best Local Similarity 13.5%; Pred. No. 0.084;  
Matches 96; Conservative 301; Mismatches 308; Indels 8; Gaps 3;

ORIGIN

Query Match 5.4%; Score 55.6; DB 9; Length 125020;  
Best Local Similarity 13.5%; Pred. No. 0.084;  
Matches 96; Conservative 301; Mismatches 308; Indels 8; Gaps 3;

214 GCGAGCTGAGCAGCATGTGTTCCAGATGTGGGAATCGAGAGAGCCCGGCACAGGCC 273  
17564 GYMSKTCYSAGKSRMYCMRMSKSSMSMSMASVCMQMGAGYRKSRSAGWGA 17505  
274 GTGCAGGAGAACCCAGAGGCTGTAGCCCGGCACATGCTCAGGCTGTGCTCC 333  
17504 GMRSSGKRSSTGKRSKTSYSTGSRSMKKKGYKYSRGMKKKTCYCMKKYKY 17445  
334 TGGCAGCAGAGCCCTACTGCTGAACGAGCAGAGATCTGAGCCCGGAGAGGCTCAG 393  
17444 RKTSMCMWYMSWGWKRYRCCKMKKCTGYRGMSSKSGYKSMRGMSSYSTSKMS 17385  
394 GAACTTCGTGAACATCTAGCAAGTCGGGCTGG-GGTGGCCAGTTAGACAGATGT 452  
17384 CMGYSMMKCMKYMSYKKRSMRSMSSKMRGAYARGCYSSSMSTRBRKCYSKSY 17325  
453 AGGACCCCTGTGA-----CTCAGAAATGGCAGCTCTTTGGCCAGAGGGCCAGCGTG 507  
17324 KKGKRMKMGKMRGSKYMSKMRSSMSKCYTSKYSKGRKRSKMGSTYSKAS 17265  
508 TGTCCGGGCTGGTACTCAGAAAGGTCACTGGGGGCTTCCACTAGACCCCGGCTG 567  
17264 SMRMAAGCTYGSYSNNRNRNMKGTCNMYRBSRAMNNGNAAAAGCTTCCANTN 17205  
568 GACATCTGTAGAGCCCGCAGGCTCGAGGAGCAGCTGAGCCCATGAGAGAGGCCAG 627  
17204 GGGGAAAAAGCGGASBASCYGRMSKSCYRSRTRRCMSKCRGRSGKSMWMTG 17145  
628 TTCTCTCTTAAGGATATGCTGTAGATGAGGAGAACAGAACGCCGAGGAGACTAA 687  
17144 RSGKTSYSAGSRGCGYCMGWRKGRKCMSSKMKMYKSYRRKRMTCMCMCYSMT 17085  
688 CCCAGATCCAGCCCGGCTCATCCCGTGTGCTCAAGCAATATCTTCTCTCTC 747  
17084 AATCCKSMCMCKSCCGCTGMSMSYS--YSKYSWMSKSYMSRYSRSTSKMRS 17027  
748 TGAGCTCTGCTGAGCCTTACAGAGGCTCAATGAGGGGGGTGAGAGAGCCAGACGCTG 807

Db 17026 KKMGTGGRYKGGGSKYKGGSGYKGGSWGCKGKSTRRSSAKKSKYKMSCCAYKS 16967

Qy 808 AAGCCTTTTACATTCCTCGGGGTAGAGCCCTTCCAAATGCTGTGATCTGCTC 867

Db 16966 MSYSKYSKMMCMGYYICWGGKCMSSCMSSMRKSKSGSWAGMWDKSVSTDSK 16907

Qy 868 ACTGCTGTGTAGGGGTCCCAACGAGGCTCAGTAGTGCGTGTGAGCTGCT 920

Db 16906 MBSRSBSKYSKMSMRGKYYCTDSMKMYTCGATGCGCTACTACTT 16854

RESULT 14  
AC021035 173791 bp DNA linear PRI 20-APR-2002

LOCUS Homo sapiens chromosome 10 clone RP11-381K7, complete sequence.

DEFINITION AC021035

ACCESSION AC021035.8 GI:20219064

VERSION HNG.

KEYWORDS Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 173791)

AUTHORS Smith,D.R.

JOURNAL Genome Therapeutics Corporation Sequencing Center: Human Genome

REFERENCE 2 (bases 1 to 173791)

AUTHORS Smith,D.R.

JOURNAL Direct Submission

REFERENCE 3 (bases 1 to 173791)

AUTHORS Smith,D.R.

JOURNAL Submitted (29-JAN-2002) Genome Therapeutics Corporation, 100 Beaver

REFERENCE 4 (bases 1 to 173791)

AUTHORS Smith,D.R.

JOURNAL Direct Submission

COMMENT Submitted (20-APR-2002) Genome Therapeutics Corporation, 100 Beaver

FEATURES On Apr 20, 2002 this sequence version replaced gi:18390167.

source Location/Qualifiers

1..173791

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="10"

/clone="RP11-381K7"

/clone\_1fb="RP11-381K7"

ORIGIN

Query Match 5.3%; Score 54.8; DB 9; Length 173791;  
Best Local Similarity 55.9%; Pred. No. 0.11; 82; Indels 0; Gaps 0;  
Matches 104; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

45 TGTGCTGAGAGGCAATTCAGTCCATTTTTCAGATGAGAGTTGAGGCCCAAGAAAC 104

Db 98700 TGAAGTGGAGAGGCTTTGTTCTTCCATTTTACAGAGAGAGAAAGAGGCTCAGGAGCT 98759

Qy 105 TAAATATCTGTGTAGGACCAACAGCTTGAAGAAAGCAGAGCCCGGAGCCCTGTGT 164

Db 98760 TAAATATCTGTGTAGGACCAACAGCTTGAAGAAAGCAGAGCCCGGAGCCCTGTGT 98819

Qy 165 GTGTGACAGCCCGGAGCCCGGAGCTGTATGCGGGGCTCGGAGCCAGAGAGCTGAG 224

Db 98820 CTGGCTTACCGAGAGCTTGTGTTGGGTGTGTGTTCTTACACCCAGAGGCTTGTGCTG 98879

Qy 225 CAGCAT 230

Db 98880 GGCAT 98885



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: June 29, 2005, 12:58:18 ; Search time 675 Seconds  
(without alignments)  
9103.247 Million cell updates/sec

Title: US-10-617-443B-1\_COPY\_1\_1038

Perfect score: 1038

Sequence: 1 cagsgactcgaagttcgcag.....ggccctcccgccagcgag 1038

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : N\_Geneseq\_16Dec04:\*

1: geneseq1980s:\*  
2: geneseq1990s:\*  
3: geneseq2000s:\*  
4: geneseq2001as:\*  
5: geneseq2001bs:\*  
6: geneseq2002as:\*  
7: geneseq2002bs:\*  
8: geneseq2003as:\*  
9: geneseq2003bs:\*  
10: geneseq2003cs:\*  
11: geneseq2003ds:\*  
12: geneseq2004as:\*  
13: geneseq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1038	100.0	3006	12 ADJ11355	Adj11355 Human PRS
2	55.8	5.4	111282	6 AAL44261	AAL44261 Human pho
3	55.8	5.4	111282	6 ABSS5190	ABSS5190 Genomic D
4	52.4	5.0	7086	4 AAL03024	AAL03024 Human rep
5	52.4	5.0	7086	4 AAL03024	AAL03024 Human rep
6	52.4	5.0	7086	4 AAL03025	AAL03025 Human rep
7	52.4	5.0	7086	4 AAL03025	AAL03025 Human rep
8	49.6	4.8	532	4 AAS32618	AAS32618 Human gen
9	49.6	4.8	532	4 AAS32618	AAS32618 Human gen
10	49.2	4.7	414	4 AAS32311	AAS32311 Rice gene
11	49.2	4.7	7353	6 AAS32311	AAS32311 Human CDN
12	48.8	4.7	2000	8 ABA171938	ABA171938 Human G-P
13	48.8	4.6	29163	4 AAL05121	AAL05121 Human rep
14	48.2	4.6	29163	4 ABL98013	ABL98013 Human tes
15	47.2	4.5	52661	9 ADA02876	ADA02876 Human DPT
16	47.2	4.5	52661	10 ADA02876	ADA02876 Human DPT
17	47.2	4.5	52661	10 ADA02876	ADA02876 Human DPT
18	47.2	4.5	52661	12 ADA02876	ADA02876 Human DPT
19	47.2	4.5	74822	6 ABL10752	ABL10752 Human bre
20	47.2	4.5	99957	9 ADA03032	ADA03032 Human mCG

## ALIGNMENTS

RESULT 1	ID	ADJ11355	standard; cDNA; 3006 BP.
XX	ADJ11355	20-MAY-2004 (first entry)	
XX	AC	ADJ11355	
XX	DT	20-MAY-2004 (first entry)	
XX	DE	Human PRS11-L cDNA encoding a serine protease Segid 1.	
XX	XX	human; PRS11-L; gene; ss; S2 serine protease; S2/HTRA;	
XX	KW	cellular physiology; apoptosis; gene therapy; osteoarthritis; stress;	
XX	KM	apoptotic; osteopathic; antiarthritic; tranquiliser.	
XX	OS	Homo sapiens.	
XX	FN	Key	Location/Qualifiers
XX	FT	CDS	1011..2015
XX	FT	FT	/*tag= a
XX	XX	XX	/product= "PRS11-L protein"
XX	PN	US2004005659-A1.	
XX	PD	08-JAN-2004.	
XX	PF	03-JUL-2002; 2002US-00189099.	
XX	PR	03-JUL-2002; 2002US-00189099.	
XX	PA	(DARR/) DARROW A L.	
XX	PA	(OLJ/) OL J.	
XX	PA	(CHEN/) CHEN C.	
XX	PA	(ANDR/) ANDRADE-GORDON P.	
XX	PI	Darrow AL, OL J, Chen C, Andrade-Gordon P;	
XX	XX	WPI: 2004-081723/08.	
XX	DR	P-PSDB; ADJ11356.	
XX	XX	New isolated S2 serine protease nucleic acids and polypeptides, useful	
XX	PT	for diagnosing and/or treating diseases with aberrant expression or	
XX	PT	activity the S2 serine protease, such as osteoarthritis, stress and	
XX	PT	apoptotic disorders.	

C 21	47.2	4.5	99957	10	ADB72770	ADB72770 Human CA
C 22	47.2	4.5	99957	10	ADC85512	ADC85512 Human gen
C 23	47.2	4.5	99957	12	ADM74627	ADM74627 Human car
C 24	47	4.5	33923	4	AAK67071	AAK67071 Human imm
C 25	47	4.5	261817	11	ACN45182	ACN45182 Human imm
C 26	46.6	4.5	13216	4	AAL05122	AAL05122 Human rep
C 27	46.6	4.5	13216	4	ABL98014	ABL98014 Human tes
C 28	46.6	4.5	257645	12	ADG97289	ADG97289 Human can
C 29	46.2	4.5	703	4	AAK87363	AAK87363 Human dev
C 30	46.2	4.5	2849	10	ADB62282	ADB62282 Human gen
C 31	46.2	4.5	63686	11	ACN44158	ACN44158 Human gen
C 32	45.4	4.4	243428	12	ADP51132	ADP51132 Human P-R
C 33	45	4.3	110000	12	ADQ44435_3	ADQ44435 (4 of
C 34	45	4.3	110000	13	ABD32806_5	ABD32806 (6 of
C 35	44.6	4.3	1741	4	AAH17445	AAH17445 Human CDN
C 36	44.6	4.3	1762	4	AAK87021	AAK87021 Human imm
C 37	44.6	4.3	1762	5	ABA19605	ABA19605 Human imm
C 38	44.6	4.3	1815	4	AAK87020	AAK87020 Human imm
C 39	44.6	4.3	1815	5	ABA19604	ABA19604 Human imm
C 40	44.6	4.3	57130	8	ABO77243	ABO77243 Human MAR
C 41	44.4	4.3	591	4	AAH09370	AAH09370 Human CDN
C 42	44.4	4.3	79597	10	ADL13866	ADL13866 Osteoarth
C 43	43.8	4.2	62118	11	ACN44566	ACN44566 Human gen
C 44	43.8	4.2	68001	12	ADJ54477	ADJ54477 Human B7-
C 45	43.8	4.2	223556	11	ACN44110	ACN44110 Human gen

XX Claim 2; SEQ ID NO 1; 28pp; English.

CC This invention relates to a novel isolated nucleic acid molecule  
CC identified as PRS511-L that encodes an S2 serine protease. Specifically,  
CC it refers to members of the S2/HtrA serine protease family, such that it  
CC plays a role in cellular physiology and apoptosis. The present invention  
CC provides agonists, antagonists, antibodies and recombinant expression  
CC vectors useful in methods of treatment, or detection and diagnosis of  
CC diseases associated with the aberrant expression or activity of the S2  
CC serine protease, PRS511-L. Accordingly, compositions described herein can  
CC be used via gene therapy routes to treat osteoarthritis, stress and  
CC apoptotic disorders. As such, they exhibit osteoparhnic, antiarthritic and  
CC tranquilliser activities. This polynucleotide sequence is the human PRS511  
CC -L cDNA sequence of the invention.

XX Sequence 3006 BP; 615 A; 918 C; 911 G; 562 T; 0 U; 0 Other;

Query Match 100.0%; Score 1038; DB 12; Length 3006;  
Best Local Similarity 100.0%; Pred. No. 8,1e-260; Mismatches 0; Indels 0; Gaps 0;  
Matches 1038; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGGAGCTGAAATTGACATCCTCCACATCAGTCCACAGATGTGTAGAGGGCAT 60  
DB 1 CAGGAGCTGAAATTGACATCCTCCACATCAGTCCACAGATGTGTAGAGGGCAT 60  
QY 61 ATTCAATCCATTTTTCAGATGAGAGTTGAGGCCCAAGAACGTAAATCTCTCA 120  
DB 61 ATTCAATCCATTTTTCAGATGAGAGTTGAGGCCCAAGAACGTAAATCTCTCA 120  
QY 121 GGCACACAGCTAGAAAGAGAGCCAGCCGAAACCCCTGGTGTGTAGAGGCCAGC 180  
DB 121 GGCACACAGCTAGAAAGAGAGCCAGCCGAAACCCCTGGTGTGTAGAGGCCAGC 180  
QY 181 CCAATTGCTCATTTGCGGGCTCGGAGCCACGACGAGGCTGAGACATGTGTTCCAGA 240  
DB 181 CCAATTGCTCATTTGCGGGCTCGGAGCCACGACGAGGCTGAGACATGTGTTCCAGA 240  
QY 241 TGGTGGGAATCTGAGAGAGCCCGGACAGGCCCTGTCAAGGAAACCCCGAGGGCTGTAAGGC 300  
DB 241 TGGTGGGAATCTGAGAGAGCCCGGACAGGCCCTGTCAAGGAAACCCCGAGGGCTGTAAGGC 300  
QY 301 CCCGTGCACTGATGCTCAAGGCTGTGTGCTCTGTGACGACAGCCCTTACTGTGACG 360  
DB 301 CCCGTGCACTGATGCTCAAGGCTGTGTGCTCTGTGACGACAGCCCTTACTGTGACG 360  
QY 361 GCAGCAGAAATCTGAGCCCGGAGAGGGTCCAGGAAAGTTCTGTAAACCATCTAGCAAGTCG 420  
DB 361 GCAGCAGAAATCTGAGCCCGGAGAGGGTCCAGGAAAGTTCTGTAAACCATCTAGCAAGTCG 420  
QY 421 GGGTGGGGTGTGGCCAAATTGACACAGATGTAGGGCCCTGTGGAATCTAGAAATTTGGCAG 480  
DB 421 GGGTGGGGTGTGGCCAAATTGACACAGATGTAGGGCCCTGTGGAATCTAGAAATTTGGCAG 480  
QY 481 CTCTTTTGGCCAGAGGGGCAAGCTGTGTCCGGGCTGTGAGTCTCAAGAGGGTCACT 540  
DB 481 CTCTTTTGGCCAGAGGGGCAAGCTGTGTCCGGGCTGTGAGTCTCAAGAGGGTCACT 540  
QY 541 GGGGGGTCTTCACTACCCCGGCTGTGACATCTGTAGAGCCCGAGGGCTCGAGGGAGCC 600  
DB 541 GGGGGGTCTTCACTACCCCGGCTGTGACATCTGTAGAGCCCGAGGGCTCGAGGGAGCC 600  
QY 601 AGCTGAGAGCCATGAGAGAGAGGGGCAATTCTCTCTGTAAAGGTATTTGCTGATGAG 660  
DB 601 AGCTGAGAGCCATGAGAGAGAGGGGCAATTCTCTCTGTAAAGGTATTTGCTGATGAG 660  
QY 661 GGAACAGACAGAGCCAGAGGGGACTAACCCGAGATCCAGGCCCGGGCTCACTCCCGTGTG 720  
DB 661 GGAACAGACAGAGCCAGAGGGGACTAACCCGAGATCCAGGCCCGGGCTCACTCCCGTGTG 720  
QY 721 GCTCAGGAGAAATCTCTCTCTCTGAGCTCTCTGAGCCCTGAGCCCTAGAGAGGTTCCAGTGTG 780  
DB 721 GCTCAGGAGAAATCTCTCTCTCTGAGCTCTCTGAGCCCTGAGCCCTAGAGAGGTTCCAGTGTG 780

RESULT 2  
AAL44261/C  
ID AAL44261 standard; DNA; 111282 BP.

QY 781 AGGGGGTGAAGAACCCAGACAGTGAAGCCTTTTAACCATCTCGGGGTGAGCGAGC 840  
DB 781 AGGGGGTGAAGAACCCAGACAGTGAAGCCTTTTAACCATCTCGGGGTGAGCGAGC 840  
QY 841 CCTTCCCAATGCTGTGTGTCTGACATGCTGTGTGTGTAGAGGGATCCCAACGGGCTC 900  
DB 841 CCTTCCCAATGCTGTGTGTGTCTGACATGCTGTGTGTGTAGAGGGATCCCAACGGGCTC 900  
QY 901 AGTGTGGCTGAGGCTGCTCTGAACTGTGAGACAGGGGTCTCAGAGAAAGCTCTCTCTCC 960  
DB 901 AGTGTGGCTGAGGCTGCTCTGAACTGTGAGACAGGGGTCTCAGAGAAAGCTCTCTCTCC 960  
QY 961 TGGCCACTGGGAGATATGGCTCTGAGAGCTGGAGAGCATGTGATCTCATGTATGACCTGG 1020  
DB 961 TGGCCACTGGGAGATATGGCTCTGAGAGCTGGAGAGCATGTGATCTCATGTATGACCTGG 1020  
QY 1021 CCTTCCCGCAGCGCAG 1038  
DB 1021 CCTTCCCGCAGCGCAG 1038

AC AAL44261;  
XX 24-OCT-2002 (first entry)  
XX Human phosphodiesterase protein (PDE) gene sequence.  
DE Human; gene; ds; gene therapy; vaccine; protein therapy;  
KW phosphodiesterase protein; PDE; PDE2A; asthma;  
KW chronic obstructive pulmonary disease; rheumatoid arthritis; atopy.  
XX Homo sapiens.

XX Key  
FH location/Qualifiers  
FT replace(254, A)  
FT /tag= a  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(1997, T)  
FT /tag= b  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(2553, A)  
FT /tag= c  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(2986, A)  
FT /tag= d  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(3702, A)  
FT /tag= e  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(3710, G)  
FT /tag= f  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(4661, T)  
FT /tag= g  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(4734, A)  
FT /tag= h  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(5649, G)  
FT /tag= i  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(5957, G)  
FT /tag= j  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(6060, C)  
FT /tag= k  
FT /standard\_name= "single nucleotide polymorphism"  
FT replace(6133, G)



```
FT /tag= l
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6692, G/C/T)
FT /tag= m
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6911, G)
FT /tag= n
FT /standard_name= "Single nucleotide polymorphism"
FT replace(7541, T)
FT /tag= o
FT /standard_name= "Single nucleotide polymorphism"
FT replace(8173, T/C/A)
FT /tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT replace(8694, G/T/C)
FT /tag= q
FT /standard_name= "Single nucleotide polymorphism"
FT replace(9072, G)
FT /tag= r
FT /standard_name= "Single nucleotide polymorphism"
FT replace(9426, G/C/T)
FT /tag= s
FT /standard_name= "Single nucleotide polymorphism"
FT replace(9573, G)
FT /tag= t
FT /standard_name= "Single nucleotide polymorphism"
FT replace(9826, T)
FT /tag= u
FT /standard_name= "Single nucleotide polymorphism"
FT replace(10134, T/A/G)
FT /tag= v
FT /standard_name= "Single nucleotide polymorphism"
FT replace(11014, T)
FT /tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT replace(12390, A)
FT /tag= x
FT /standard_name= "Single nucleotide polymorphism"
FT replace(14701, T/A/G)
FT /tag= y
FT /standard_name= "Single nucleotide polymorphism"
FT replace(15679, T)
FT /tag= z
FT /standard_name= "Single nucleotide polymorphism"
FT replace(18606, T)
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism"
FT replace(19070, G)
FT /tag= ab
FT /standard_name= "Single nucleotide polymorphism"
FT replace(19470, G)
FT /tag= ac
FT /standard_name= "Single nucleotide polymorphism"
FT replace(19611, A)
FT /tag= ad
FT /standard_name= "Single nucleotide polymorphism"
FT replace(20641, C)
FT /tag= ae
FT /standard_name= "Single nucleotide polymorphism"
FT replace(20642, C)
FT /tag= af
FT /standard_name= "Single nucleotide polymorphism"
FT replace(21036, T)
FT /tag= ag
FT /standard_name= "Single nucleotide polymorphism"
FT replace(21871, C)
FT /tag= ah
FT /standard_name= "Single nucleotide polymorphism"
FT replace(22907, A)
FT /tag= ai
FT /standard_name= "Single nucleotide polymorphism"
FT replace(23722, G)
FT /tag= aj
FT /standard_name= "Single nucleotide polymorphism"
FT replace(25121, C)
FT /tag= ak
FT /standard_name= "Single nucleotide polymorphism"
FT replace(24553, C)
FT /tag= al
FT /standard_name= "Single nucleotide polymorphism"
FT replace(25917, C)
FT /tag= am
FT /standard_name= "Single nucleotide polymorphism"
FT replace(27525, A)
FT /tag= an
FT /standard_name= "Single nucleotide polymorphism"
FT replace(27625, C/A)
FT /tag= ao
FT /standard_name= "Single nucleotide polymorphism"
FT replace(27833, T)
FT /tag= ap
FT /standard_name= "Single nucleotide polymorphism"
FT replace(27852, G)
FT /tag= aq
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28478, A)
FT /tag= ar
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28514, C/G)
FT /tag= as
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28702, C)
FT /tag= at
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28859, C/G/T)
FT /tag= au
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28960, G)
FT /tag= av
FT /standard_name= "Single nucleotide polymorphism"
FT replace(29030, G)
FT /tag= aw
FT /standard_name= "Single nucleotide polymorphism"
FT replace(29348, C)
FT /tag= ax
FT /standard_name= "Single nucleotide polymorphism"
FT replace(29973, C/G/A)
FT /tag= ay
FT /standard_name= "Single nucleotide polymorphism"
FT replace(30153, T)
FT /tag= az
FT /standard_name= "Single nucleotide polymorphism"
FT replace(30389, C)
FT /tag= ba
FT /standard_name= "Single nucleotide polymorphism"
FT replace(30581, G)
FT /tag= bb
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31147, A)
FT /tag= bc
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31224, T)
FT /tag= bd
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31735, A/T/C)
FT /tag= be
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31798, C/A/T)
FT /tag= bf
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31994, G/T)
FT /tag= bg
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32324, T)
FT /tag= bh
FT /standard_name= "Single nucleotide polymorphism"
```



```
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(20642,A)
FT /tag= ak
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(21036,T)
FT /tag= al
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(21871,C)
FT /tag= am
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(22907,A)
FT /tag= an
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(23722,G)
FT /tag= ao
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(24121,C)
FT /tag= ap
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(24553,C)
FT /tag= aq
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(25917,C)
FT /tag= ar
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(26572,.26574,CG)
FT /tag= as
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(27525,A)
FT /tag= at
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(27625,Y)
FT /tag= au
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(27833,C)
FT /tag= av
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(27852,G)
FT /tag= aw
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(28478,C)
FT /tag= ax
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(28514,Y)
FT /tag= ay
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(28702,C)
FT /tag= az
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(28859,H)
FT /tag= ba
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(28960,G)
FT /tag= bb
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(29030,G)
FT /tag= bc
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(29348,C)
FT /tag= bd
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(29973,H)
FT /tag= be
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(30153,T)
FT /tag= bf
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(30389,C)
FT /tag= bg
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(30581,C)
FT /tag= bh
FT /standard_name="Single nucleotide polymorphism (SNP)"
```

```
FT variation
FT /tag= bi
FT /standard_name="Single nucleotide polymorphism (SNP)"
FT replace(31224,C)
FT /tag= bj
FT /standard_name="Single nucleotide polymorphism (SNP)"

Query Match
Best Local Similarity 70.1%; Pred. No. 0.001; Indels 1; Gaps 1;
Matches 89; Conservative 0; Mismatches 37;

Qy 67 TCCCATTTTCAGATGAGAGTTGAGGCCGAGAGACGTAACTGTCGTGAGGCCAC 126
Db 93162 TCCCATTTTCAGGTGTGAAACTGAGGCCGAGAGGTTAGTACTTGTCTCAAGGTAC 93103

Qy 127 ACAGCTGAAAGCAGCCAGCC-CAGCCGAACCTTGCTGTGACAGCCCACT 185
Db 93102 ACAGCTCAAAAGCAGCAAGCCGAACTGAAACCCAGAGTCTGCACCTCCGACAG 93043

Qy 186 TGCCTCAT 192
Db 93042 CCATCTT 93036

RESULT 4
AAL03024
ID AAL03024 standard; DNA; 7086 BP.
AC AAL03024;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 5712.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN W0200155320-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001MO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216477P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
```

PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226719P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231422P.  
PR 08-SEP-2000; 2000US-0231423P.  
PR 08-SEP-2000; 2000US-0231444P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 13-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 01-NOV-2000; 2000US-0244647P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0255719P.  
PR 06-DEC-2000; 2000US-0251472P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-02559678P.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
DR  
XX  
XX  
PT Isolated nucleic acid molecule encoding a reproductive system antigen is  
XX used in preventing, treating or ameliorating a medical condition.  
PS Disclosure; SEQ ID NO 5712; 1297pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention  
XX  
SQ Sequence 7086 BP; 1264 A; 2050 C; 1979 G; 1793 T; 0 U; 0 Other;  
Query Match 5.0%; Score 52.4; DB 4; Length 7086;  
Best.Local Similarity 69.6%; Pred.No. 0.0036;  
Matches 71; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 68 CCCATTTTCAGATGAGAGTTGAGGCCGAGAAAGTAAATCTGTTGTAAGCCACA 127  
Db 4323 CTCATTTTATAGATGAAACATGCGGCGCTAGAGAAAGATTAATCTCTCAAGCCACA 4382  
QY 128 CAGCTAGAAAGCAGCAGGCCGAGCCGAAACCCCTGCTGTG 169  
Db 4383 CAGCTAGAAAGTACAGAGCTTATGCACTGACGTAGGGTTTGTG 4424  
RESULT 5  
ID ABA17415 standard; DNA; 7086 BP.  
XX  
AC ABA17415;



PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0251989P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251899P.  
PR 08-DEC-2000; 2000US-0251907P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
XX PT useful for preventing, diagnosing and/or treating nervous system cancers  
XX PT and metastases.  
XX  
XX Disclousre; SEQ ID NO 9746; 1701bp + Sequence listing; English.  
XX  
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins  
XX CC (AB14678-AB18001) useful for preventing, treating or ameliorating  
XX CC medical conditions e.g. by protein or gene therapy. The genes are  
XX CC isolated from a range of human tissues disclosed in the specification.  
XX CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
XX CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
XX CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
XX CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
XX CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic  
XX CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
XX CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
XX CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
XX CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
XX CC infectious diseases such as viral, bacterial, fungal and parasitic  
XX CC infections. Note: The sequence data for this patent did not form part of  
XX CC the printed specification, but was obtained in electronic format directly  
XX CC from WIPO at ftp.wipo.int/pub/publshd\_pct\_sequences  
XX  
SQ Sequence 7086 BP; 1264 A; 2050 C; 1979 G; 1793 T; 0 U; 0 Other;  
Query Match 5.0%; Score 52.4; DB 5; Length 7086;  
Best Local Similarity 69.6%; Pred. No. 0.0036;  
Matches 71; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 68 CCCATTTTCAGATGAGAGTTGAGGCCCGAGAGACGTAAGTATCTGTCTGAGGCCACA 127  
Db 4323 CTCATTTTATAGTGAACACACGTGGGCTTAGAGAAAGATATCTCTCCAAAGGCCACA 4382  
QY 128 CAGCTAGAAAGCAGCGAGGCCCGAACCCCTGTGTGTG 169  
Db 4383 CAGCTAGAAAGTACAGAGCTTATGCCACGTACAGGTTGTG 4424  
RESULT 6  
AA103025  
ID AA103025 standard; DNA; 7388 BP.  
XX  
XX AA103025;  
XX  
XX 21-NOV-2001 (first entry)  
XX  
XX Human reproductive system related antigen DNA SEQ ID NO: 5713.  
XX  
XX Human reproductive system related antigen; reproductive system disorder;  
XX cancer; gene therapy; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200155320-A2.  
PN

XX  
PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001339.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214866P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216800P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226868P.  
XX 23-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0228928P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 05-SEP-2000; 2000US-0229509P.  
XX 05-SEP-2000; 2000US-0229513P.  
XX 06-SEP-2000; 2000US-0230437P.  
XX 06-SEP-2000; 2000US-0230438P.  
XX 08-SEP-2000; 2000US-0231242P.  
XX 08-SEP-2000; 2000US-0231243P.  
XX 08-SEP-2000; 2000US-0231244P.  
XX 08-SEP-2000; 2000US-0231413P.  
XX 08-SEP-2000; 2000US-0231414P.  
XX 08-SEP-2000; 2000US-0232081P.  
XX 08-SEP-2000; 2000US-0232081P.  
XX 12-SEP-2000; 2000US-0233968P.  
XX 14-SEP-2000; 2000US-0233977P.  
XX 14-SEP-2000; 2000US-0233988P.  
XX 14-SEP-2000; 2000US-0233999P.  
XX 14-SEP-2000; 2000US-0234000P.  
XX 14-SEP-2000; 2000US-0234011P.  
XX 14-SEP-2000; 2000US-0234063P.  
XX 14-SEP-2000; 2000US-0234064P.  
XX 14-SEP-2000; 2000US-0234065P.  
XX 21-SEP-2000; 2000US-0234223P.  
XX 21-SEP-2000; 2000US-0234274P.  
XX 25-SEP-2000; 2000US-0234997P.  
XX 25-SEP-2000; 2000US-0234998P.  
XX 26-SEP-2000; 2000US-0235484P.  
XX 27-SEP-2000; 2000US-0235834P.  
XX 27-SEP-2000; 2000US-0235836P.  
XX 29-SEP-2000; 2000US-0236327P.  
PR

PR	29-SEP-2000;	2000US-0236367P.
PR	29-SEP-2000;	2000US-0236368P.
PR	29-SEP-2000;	2000US-0226358P.
PR	29-SEP-2000;	2000US-0236370P.
PR	02-OCT-2000;	2000US-0236802P.
PR	02-OCT-2000;	2000US-0237037P.
PR	02-OCT-2000;	2000US-0237038P.
PR	02-OCT-2000;	2000US-0237039P.
PR	02-OCT-2000;	2000US-0237040P.
PR	13-OCT-2000;	2000US-0239935P.
PR	13-OCT-2000;	2000US-0239937P.
PR	20-OCT-2000;	2000US-0240960P.
PR	20-OCT-2000;	2000US-0241221P.
PR	20-OCT-2000;	2000US-0241785P.
PR	20-OCT-2000;	2000US-0241786P.
PR	20-OCT-2000;	2000US-0241787P.
PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241825P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249246P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249257P.
PR	17-NOV-2000;	2000US-0249259P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251988P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Barash SC, Ruben SM;	
XX		

[illegible]

PR	14-AUG-2000	2000US-02244519P
PR	14-AUG-2000	2000US-0225213P
PR	14-AUG-2000	2000US-0225214P
PR	14-AUG-2000	2000US-0225266P
PR	14-AUG-2000	2000US-0225267P
PR	14-AUG-2000	2000US-0225268P
PR	14-AUG-2000	2000US-0225270P
PR	14-AUG-2000	2000US-0225477P
PR	14-AUG-2000	2000US-0225757P
PR	14-AUG-2000	2000US-0225758P
PR	14-AUG-2000	2000US-0225759P
PR	18-AUG-2000	2000US-0226279P
PR	22-AUG-2000	2000US-0226681P
PR	22-AUG-2000	2000US-0226688P
PR	22-AUG-2000	2000US-0227182P
PR	30-AUG-2000	2000US-0227009P
PR	30-AUG-2000	2000US-0228924P
PR	01-SEP-2000	2000US-0229287P
PR	01-SEP-2000	2000US-0229343P
PR	01-SEP-2000	2000US-0229344P
PR	01-SEP-2000	2000US-0229345P
PR	05-SEP-2000	2000US-0229509P
PR	05-SEP-2000	2000US-0229513P
PR	06-SEP-2000	2000US-0230437P
PR	06-SEP-2000	2000US-0230438P
PR	08-SEP-2000	2000US-0231242P
PR	08-SEP-2000	2000US-0231243P
PR	08-SEP-2000	2000US-0231244P
PR	08-SEP-2000	2000US-0231413P
PR	08-SEP-2000	2000US-0231414P
PR	08-SEP-2000	2000US-0232080P
PR	12-SEP-2000	2000US-0232081P
PR	12-SEP-2000	2000US-0231968P
PR	14-SEP-2000	2000US-0232397P
PR	14-SEP-2000	2000US-0232398P
PR	14-SEP-2000	2000US-0232399P
PR	14-SEP-2000	2000US-0232400P
PR	14-SEP-2000	2000US-0232401P
PR	14-SEP-2000	2000US-0232403P
PR	14-SEP-2000	2000US-0232404P
PR	14-SEP-2000	2000US-0232365P
PR	21-SEP-2000	2000US-0234223P
PR	21-SEP-2000	2000US-0234274P
PR	25-SEP-2000	2000US-0234997P
PR	25-SEP-2000	2000US-0234998P
PR	26-SEP-2000	2000US-0235484P
PR	27-SEP-2000	2000US-0235834P
PR	27-SEP-2000	2000US-0235835P
PR	29-SEP-2000	2000US-0236377P
PR	29-SEP-2000	2000US-0236378P
PR	29-SEP-2000	2000US-0236368P
PR	29-SEP-2000	2000US-0236369P
PR	29-SEP-2000	2000US-0236370P
PR	02-OCT-2000	2000US-0237037P
PR	02-OCT-2000	2000US-0237038P
PR	02-OCT-2000	2000US-0237039P
PR	02-OCT-2000	2000US-0237040P
PR	13-OCT-2000	2000US-0239395P
PR	13-OCT-2000	2000US-0239397P
PR	20-OCT-2000	2000US-0241785P
PR	20-OCT-2000	2000US-0241786P
PR	20-OCT-2000	2000US-0241787P
PR	20-OCT-2000	2000US-0241808P
PR	20-OCT-2000	2000US-0241809P
PR	20-OCT-2000	2000US-0241866P
PR	01-NOV-2000	2000US-0244221P
PR	01-NOV-2000	2000US-0244617P
PR	08-NOV-2000	2000US-0246475P
PR	08-NOV-2000	2000US-0246476P
PR	08-NOV-2000	2000US-0246477P

PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250391P.
PR	01-DEC-2000;	2000US-0251160P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-541565/60.	
DR		
PT	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating nervous system cancers	
PT	and metastases.	
XX		
XX	Disclosure; SEQ ID NO 9747; 1701bp + Sequence Listing; English.	
XX		
XX	The invention relates to novel genes (ABN11004-ABN21534) and proteins	
CC	(ABN1678-ABN18001) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful in	
CC	the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and	
CC	ovarian cancer and other cancers of the adrenal gland, bone, bone marrow	
CC	breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune	
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic	
CC	anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,	
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)	
CC	cardiovascular disorders such as myocardial ischaemias; (d) wound healing	
CC	; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)	
CC	infectious diseases such as viral, bacterial, fungal and parasitic	
CC	infections. Note: The sequence data for this patent did not form part of	
CC	the printed specification, but was obtained in electronic format directly	



CC from WIP0 at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 7388 BP, 1347 A; 2127 C; 2078 G; 1836 T; 0 U; 0 Other;  
SQ  
Query Match 5.0%; Score 52.4; DB 5; Length 7388;  
Best Local Similarity 69.6%; Freq. No. 0.0036;  
Matches 71; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY  
68 CCCATTTTCAGATGAGAGTTGAGCCCGACAGAACGTAATCTGTGAGGCCACA 127  
Db 4327 CTCATTATTATGATGAAACACTGGGGCCTAGAGAAAGATTAATCTTCCAGAGCCACA 4386  
QY  
128 CAGCTAGAAAGCAGCCAGGCCCGAACCCCTGTGTGTG 169  
Db 4387 CAGCTAGAAAGTACAGAGCTTATGCCACGTCAGGTTTGTG 4428  
RESULT 8  
AAS32618/c  
ID AAS32618 standard; DNA; 532 BP.  
XX  
AC AAS32618;  
XX  
DT 17-DEC-2001 (first entry)  
XX  
DE Human genomic DNA for novel endocrine antigen, SEQ ID No 572.  
XX  
KM Human; endocrine antigen; ds; cytosolic; antifertility; antidiabetic;  
KM thyroid-active; adrenal-active; androgenic; gastric; gene therapy;  
KM antisense-therapy; antibody; endocrine disorder; hormone imbalance;  
KM reproductive disorder; endocrine cancer; pancreatic disorder;  
KM diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;  
KM hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.  
XX  
OS Homo sapiens.  
XX  
PN WO200155319-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001335.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217486P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231245P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232337P.  
PR 14-SEP-2000; 2000US-0232338P.  
PR 14-SEP-2000; 2000US-0232339P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236337P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.



[illegible]



PT antibodies.  
XX  
PS Claim 4; Fig 3A-C; 72pp; English.  
XX  
CC The present sequence encodes a human G-protein coupled receptor (GPCR)  
CC protein (1). (1) can be used in gene therapy. (1) can be used for  
CC identifying a modulator of (1) by contacting (1) with an agent and  
CC determining if the agent has modulated the function or activity of (1).  
CC (1) is also useful for identifying an agent that binds to (1), by  
CC contacting (1) with an agent and assaying the contacted mixture to  
CC determine whether a complex is formed with the agent bound to (1). (1)  
CC can be used in the treatment of a disorder characterised by the absence  
CC of, inappropriate or unwanted expression of the receptor protein. (1) is  
CC located to human chromosome 1  
XX  
SQ Sequence 7353 BP; 1772 A; 1742 C; 1920 G; 1919 T; 0 U; 0 Other;  
Query Match 4.7%; Score 49.2; DB 6; Length 7353;  
Best Local Similarity 59.2%; Pred. No. 0.025;  
Matches 84; Conservative 0; Mismatches 58; Indels 0; Gaps 0;  
Qy 44 ATGTGTGAGGCGCATTTTCAATGTCCTTTTTCAGATGAGAGTTGAGCCCGAGAGAC 103  
Db 2225 ATGAGTAGATATATATTTTATCTTCATTTGCAAGATGAGAGCTGAGCTAGAAAG 2166  
Qy 104 GTAGTATCTGTCTGAGGCGCACAGCTAGAGAGCCAGCCGCGAGCCGCTGG 163  
Db 2165 TTAAAGTACTTATCCCAAGGTGCGACAGCTAGAGCGGAGGCTGGGCACTACCTGG 2106  
Qy 164 TGTGTGAGCGCCCGCAGCCAGT 185  
Db 2105 GCAGTCTGGCCCGCAGGCTCAGT 2084  
RESULT 12  
ADA71938/C  
ID ADA71938 standard; DNA; 2000 BP.  
XX  
AC ADA71938;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Rice gene, SEQ ID 5263.  
XX  
KM plant; bacterial infection; fungal infection; viral infection; rice;  
KM gene; ds.  
XX  
OS Oryza sativa.  
XX  
PN WO200300898-A1.  
XX  
PD 03-JAN-2003.  
XX  
PF 22-JUN-2001; 2001WO-IB001105.  
XX  
PR 22-JUN-2001; 2001WO-IB001105.  
XX  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
XX  
PT Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
PS Claim 27; SEQ ID NO 5263; 899pp; English.  
XX  
CC The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1

CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC the expression of the gene in an uninfected plant. In a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX  
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
Query Match 4.7%; Score 48.8; DB 8; Length 2000;  
Best Local Similarity 7.8%; Pred. No. 0.021;  
Matches 48; Conservative 303; Mismatches 265; Indels 2; Gaps 2;  
Qy 421 GGGTGGGTGTGGCCAAAGTTAGACAGATGTAGGCGCCCTGTGACTCAGAAATGGCAG 480  
Db 709 KRAKRSYRYYRRRWWKKGWTTYYRWRSCMTARRSKRRKAGASMKSCMYWRGARS 650  
Qy 481 CTCTTTGGCCAGAGGGGCGACGCTGTGTCGGGCGCTGGGTAGCTCAGAGGCTCAGCT 540  
Db 649 MWTSKYSKSAKCKKTRMTSTMTSTMTSTMTSTMTSTMTSTMTSTMTSTMTSTMTSTMT 590  
Qy 541 GGGGCTTCTTCACTACACCCCGCTGACACTGCTGTAGCCCGCAGGCTCGAGGAGCC 600  
Db 589 STRSKKRWGSGMSBMYRWKXKWRKRYRWKMKCTWRKCMCYRWGTYMTTSRBRM 530  
Qy 601 AGCTGAGCCCAAGAGAGAGGCGCACTTCTCTCTAGAGGTATGCTGTAGCATAG 660  
Db 529 YTRGKARAYTS-KRRYMWYRKYRKYCWYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 471  
Qy 661 GGAACAGACAGAGGCGGAGGAGACTAACCCGAGATCCAGCCCGGCTCACTCCGCTGTG 720  
Db 470 YSGMMWYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKY 411  
Qy 721 GCTCAGCGCAATATCTTAACCTCTCTGAGCCTCTGCGCCAGCTAGAGGCTCCAGTG 780  
Db 410 MGSYKYSRCYKN-RMYWYGMWYMYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 352  
Qy 781 AGGGGGGTGAGAAAGCCAGAGCTGAGAGCCCTTTTAACTATCTCGGGGTAGCAGC 840  
Db 351 SYKKYCTWYKWK 292  
Qy 841 CCTTCCCAATGCTGTGCTCACTGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900  
Db 291 MRTAGKMKRMSWRSRMSRMSRMSRMSRMSRMSRMSRMSRMSRMSRMSRMSRMSRMSRMS 232  
Qy 901 AGTGTGGCTGAGGCTGCTCTGAACCTGAGAGAGGCTCTCAGAAAGAGCTCTCTCC 960  
Db 231 SRRAKMMRACRMYSACRRYRTSYCCGCSYCGSSKMKYMSKSMMTSCSWSGCCCTCY 172  
Qy 961 TGCCCACTGGCATAGGCTCTGAGAGCTGAGAGCTGTGATCTCAGTATGATGACACTGG 1020  
Db 171 YGAMCWSGMSWYMGSCGCTGTGWRKRSKYSKCKKCYSCCTKYSCTGYVCKWKYK 112  
Qy 1021 CCTTCCCGCCAGGCGAG 1038  
Db 111 SYKCYCYCYWYMSYR 94  
RESULT 13  
AAL05121/C  
ID AAL05121 standard; DNA; 29163 BP.  
XX  
AC AAL05121;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen DNA SEQ ID NO: 7809.  
XX  
KM Human; reproductive system related antigen; reproductive system disorder;  
KM cancer; gene therapy; ds.  
XX

OS Homo sapiens.  
XX  
PN W0200155320-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US001339.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180638P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231142P.  
PR 08-SEP-2000; 2000US-0231143P.  
PR 08-SEP-2000; 2000US-0231144P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234597P.  
PR 25-SEP-2000; 2000US-0234598P.  
PR 26-SEP-2000; 2000US-0235484P.

PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241825P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249219P.  
PR 17-NOV-2000; 2000US-0249220P.  
PR 17-NOV-2000; 2000US-0249221P.  
PR 17-NOV-2000; 2000US-0249222P.  
PR 17-NOV-2000; 2000US-0249223P.  
PR 17-NOV-2000; 2000US-0249224P.  
PR 17-NOV-2000; 2000US-0249225P.  
PR 17-NOV-2000; 2000US-0249226P.  
PR 17-NOV-2000; 2000US-0249227P.  
PR 17-NOV-2000; 2000US-0249228P.  
PR 17-NOV-2000; 2000US-0249229P.  
PR 17-NOV-2000; 2000US-0249230P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251188P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251866P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0255678P.

XX  
PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465570/50.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
XX used in preventing, treating or ameliorating a medical condition.  
XX  
XX Disclosure; SEQ ID NO 7809; 1297pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
XX number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders,  
XX including cancer. The present sequence is a genomic sequence encoding a  
XX protein of the invention  
XX  
SQ Sequence 29163 BP; 6801 A; 7298 C; 7317 G; 7747 T; 0 U; 0 Other;  
Query Match 4.6%; Score 48.2; DB 4; Length 29163;  
Best Local Similarity 67.3%; Pred. No. 0.066;  
Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
CY 51 AGAGAGGCAATTCAGTCCATTTTTCAGATGAGAGTGAAGCCGACAGACCTAAGTA 110  
DB 8883 AGGTGAGTGTGTATTCCTCATTTTACAAATGAGGATGAGGCTCAGAGAGTGAATT 8824  
CY 111 ATCTGTGAGGCCACACAGCTAGAGAGCCAGGCCAG 151  
DB 8823 ACCAGTCCAGATTACACAGCTAATGAGCTCAGAGTCCAG 8783  
RESULT 14  
ABL98013/c  
ID ABL98013 standard; DNA; 29163 BP.  
XX  
XX ABL98013;  
AC  
XX  
XX 21-JUN-2002 (first entry)  
DT  
XX  
XX Human testicular antigen encoding DNA fragment SEQ ID NO: 2665.  
DE  
XX  
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;  
XX reproductive system disorder; urinary system disorder; gene therapy;  
XX cardiovascular disorder; respiratory disorder; neurological disorder;  
XX gastrointestinal disease; infection; cytostatic; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO20015317-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US001329.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217486P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234397P.  
PR 25-SEP-2000; 2000US-0234398P.  
PR 26-SEP-2000; 2000US-0235844P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.

XX 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-02469207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-483232/52.  
DR  
XX  
XX  
PT Nucleic acids encoding 973 human testicular antigen polypeptides, useful  
PT for preventing, diagnosing and/or treating testicular cancer.  
XX  
XX  
PS Disclosure; SEQ ID NO 2665; 766pp; English.  
XX  
CC The present invention provides the protein and coding sequences of 973  
CC human testicular antigens, and fragments of their genomic sequences. The  
CC sequences can be used in the treatment of cardiovascular, urinary system,  
CC reproductive system, immune, respiratory, neurological and  
CC gastrointestinal disorders, infections, and particularly cancer,  
CC especially testicular cancers. The present sequence is a DNA encoding a  
CC protein fragment of the invention  
XX  
SQ Sequence 29163 BP; 6801 A; 7298 C; 7317 G; 7747 T; 0 U; 0 Other;

Query Match	4.6%	Score 48.2;	DB 4;	Length 29163;
Best Local Similarity	67.3%;	Pred. No. 0.066;		
Matches 68:	Conservative	0;	Mismatches 33;	Indels 0;
				Gaps 0;

Qy 51 AGGAGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTGGAGGCCAGAGAACCTAAGTA 110

bB 8883 AGGTGAGTGTGTTATTCCCATTTTACAATGAGGAAATTGAGGCTCAGAGAGCGTGAATT 8824

```

Oy      111 ATCTGTCTGAGGCCACACAGCTAGAAAGCAGCCAGGCCAG 151
          |||||
Db      8823 ACCAGTCCAAGATTACACAGCTAATGAGCTGCAGAGTCCAG 8783

```

## RESULT 15

ID ADA02876 standard; DNA; 52661 BP.

AC ADA02876;

DT 06-NOV-2003 (first entry)

DE Human DPT carcinoma associated gene, SEQ ID NO:1394.

KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;

KW gene; ds.

OS Homo sapiens.

PN WO2003057146-A2.  
XY

PD 17-JUL-2003.  
XX

PF 26-DEC-2002; 2002WO-US041414.  
YY

PR 26-DEC-2001; 2001US-00035832.  
XX

PA (SAGR-) SAGRES DISCOVERY.  
XX

PI MORRIS DW;  
XX

WFL; 2003-38/066/33-  
XX

PT useful for preparing compositions for treating carcinomas.

Claim 1; SEQ ID NO 1394; 245pp; English.

2A The invention relates to recombinant carcinoma associated (CA) nucleic  
2B acid sequences from mouse and human (AD010182-AD03094), and to  
2C recombinant carcinoma associated proteins (CAp) encoded by them. The  
2D invention also encompasses expression vectors and host cells comprising a  
2E CA nucleic acid, a polypeptide (especially an antibody) that specifically  
2F binds to the protein, and a bioclip comprising CA nucleic acid or  
2G fragments thereof. The sequences of the invention were identified using  
2H oncogenic retroviruses, which insert into the genome of the host organism  
2I at random. Many of these do not carry transduced host oncogenes or  
2J pathogenic trans-acting viral genes, meaning that cancer incidence is a  
2K direct consequence of the effects of proviral integration into host  
2L protoconsequences. The CA nucleic acid sequences can be used to diagnose  
2M carcinoma (especially breast cancer, prostate cancer, lymphoma or  
2N leukaemia) or a propensity to carcinoma by determination of the sequence  
2O of a CA gene, or by determination of CA gene expression in particular  
2P tissues. CA nucleic acids, proteins and antibodies are also useful as  
2Q therapeutic agents and in screening and evaluating drug candidates. The  
2R present sequence represents a specifically claimed human CA nucleic acid  
2S sequence of the invention. Note: The complete sequence data for this  
2T patent did not form part of the printed specification, but was obtained  
2U in electronic format directly from WIPO at

Sequence 52661 BP; 15800 A; 11001 C; 11082 G; 14778 T; 0 U; 0 Other;

Query Match	4.5%	Score 47.2;	DB 9;	Length 5266;
Best Local Similarity	55.5%;	Pred. No. 0.14;		
Matches	91;	Conservative	0;	Mismatches 73;
			Indels	0;
			Gaps	0;

	18	CAGTCTCCACACTCAAGTTCCCACAGATGTCGTAGAAGGCATATTCAATGCCCATTTTTTC	77
QY			
Db	49680	CAGGCTTTCCACTTTGGTCTACTGCGAAGTACCCTTAGCTGSCATAGTTACATGCCCATTTTAT	49739



Oy 78 AGATGAGAGTTGAGGCCGAGAACTAAGTAATCTGTGAGGCCACACAGCTAGAAA 137  
DB 49740 AGATAGGTTCAAGGTCCCCAAATATATAGTGACTTCCCAAGTCAACAGCTAGTGG 49799  
Oy 138 GCAGCCAGGCCCGAGAACCCCTGTGTGTGACAGCCCCCAAGCC 181  
DB 49800 AAATGATGTTGAAGTGAAGTCTAGCTGTTTACAAAGCCCAAGTC 49843

Search completed: June 29, 2005, 15:39:57  
Job time : 682 secs

**THIS PAGE BLANK (USPTO)**



GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12115  
; LENGTH: 104475  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(104475)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12115

Query Match  
Best Local Similarity 5.4%; Score 55.8; DB 4; Length 104475;  
Best Local Similarity 70.1%; Pred. No. 5.4e-05;  
Matches 89; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

QY 67 TCCCATTTTCAGATGAGGAGTTGAGGCCAGAGAAAGTAATCTGTCTGAGGCCAC 126  
DB 87505 TCCCATTTTCAGATGAGGAGTTGAGGCCAGAGAGTTGAGTCTGTCTGAGGCCAC 87446  
QY 127 ACAGCTAGAAAGCAGCCAGCC-CAGCCGAACCCCTGTGTGTGAGCCCCAGCCAGT 185  
DB 87445 ACAGCTCAAAAGCAGCAAGCCGAGCTGAACCCAGTGTGTGTGAGCCCCAGCCAG 87386  
QY 186 TGCTCAT 192  
DB 87385 CCATCTT 87379

RESULT 3  
US-09-754-250-3/C  
; Sequence 3, Application US/09754250  
; Patent No. 6376225  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui et al.  
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE  
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
; TITLE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL001063  
; CURRENT APPLICATION NUMBER: US/09/754,250  
; CURRENT FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 111282  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(111282)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-754-250-3

Query Match  
Best Local Similarity 5.4%; Score 55.8; DB 3; Length 111282;  
Best Local Similarity 70.1%; Pred. No. 5.5e-05;  
Matches 89; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

QY 67 TCCCATTTTCAGATGAGGAGTTGAGGCCAGAGAAAGTAATCTGTCTGAGGCCAC 126  
DB 93162 TCCCATTTTCAGATGAGGAGTTGAGGCCAGAGAAAGTAATCTGTCTGAGGCCAC 93103  
?

QY 127 ACAGCTAGAAAGCAGCCAGCC-CAGCCGAACCCCTGTGTGTGAGCCCCAGCCAGT 185  
DB 93102 ACAGCTCAAAAGCAGCAAGCCGAGCTGAACCCAGTGTGTCTGAGGCCAGG 93043  
QY 186 TGCTCAT 192  
DB 93042 CCATCTT 93036

RESULT 4  
US-09-949-016-12822/C  
; Sequence 12822, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12822  
; LENGTH: 256171  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(256171)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12822

Query Match  
Best Local Similarity 4.8%; Score 50; DB 4; Length 256171;  
Best Local Similarity 72.2%; Pred. No. 0.0036;  
Matches 65; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 59 ATATTCAGTCCCATTTTCAGATGAGGAGTTGAGGCCAGAGAAAGTAATCTGTCT 118  
DB 189846 ATTTGTTAACCCCATTTTACAGATGAGAAATGACAGAAAGTAATCTGTCT 189787  
QY 119 GAGCCACACACAGCTAGAAAGCAGCCAGCC 148  
DB 189786 AAGTTCACACACAGCTAGTGTGAGTGCAAGCC 189757

RESULT 5  
US-09-949-016-15524/C  
; Sequence 15524, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15524  
; LENGTH: 256176

```

; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) ... (256176)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15524

```

```

Query Match
Best Local Similarity 4.8%; Score 50; DB 4; Length 256176;
Matches 65; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```

```

Qy 59 ATATTACGTCATTTTTCAGATGAGAGTGGAGCCAGAGGAGGAGTATCTGCT 118
Db 189846 ATTGTTACCCCATTTTACATGAGAAATGAGACAGAAAGATTATGATCCGTC 189787

```

```

Qy 119 GAGGCCACACAGCTAGAAAGAGCCAGGCC 148
Db 189786 AAGGTACACAGCTAGTGAAGTGCAGAACCC 189757

```

```

RESULT 6
US-09-949-016-16177
; Sequence 16177, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16177
; LENGTH: 24817
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16177

```

```

Query Match
Best Local Similarity 4.6%; Score 47.8; DB 4; Length 24817;
Matches 91; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```

```

Qy 29 ACTCAGTTCCACAGATGTGTAGAGGAGGATATTCAGTCCATTTTTCAGATGAGAGT 88
Db 7117 ACTCAGAAATAGACATAGACAGATGATCAATTAATTCATTCATGATGAGAAC 7176

```

```

Qy 89 TGAGGCCAGAGAGCTAATCTGTCTGAGGCCACAGAGTAAAGAGAGCCAGGCC 148
Db 7177 TAAGGACAGAGAGCTTAAGAAAGTTGCGAGAGGTCAACAGCTTAATTAAGTGTCTCC 7236

```

```

Qy 149 CAGCCGAAACCCCTGTGTGTGTGACGCCGCCAGCCAGCTGTCTCA 191
Db 7237 AGGCGATGGGACATGATCATCCAGCTCTGATTCATGTGTCTCA 7279

```

```

RESULT 7
US-09-621-976-14118/c
; Sequence 14118, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2

```

```

; ORGANISM: Homo sapiens
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 14118
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-14118

```

```

Query Match
Best Local Similarity 4.5%; Score 47.2; DB 4; Length 473;
Matches 81; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

```

```

Qy 30 CTCAGTTCCACAGATGTGTGTAGAGGAGTATTCAGTCCATTTTTCAGATGAGG-AGT 88
Db 194 CTCACACACCCCTGTAGTGTAGTGTGATTTTATTCATTTTACATGAGGGAAC 135

```

```

Qy 89 TGAGGCCAGAGAGCTAATCTGTCTGAGGCCACAGAGCTAGAAAGAGCCAGGCC 148
Db 134 TGAGGCACAGAGAGCTTAAGTGAATGACTTCCAGGCCACAGAGCCAGCCGCTCAGGCC 75

```

```

RESULT 8
US-09-949-016-17032
; Sequence 17032, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17032
; LENGTH: 194937
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17032

```

```

Query Match
Best Local Similarity 4.5%; Score 47; DB 4; Length 194937;
Matches 83; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

```

Qy 1 CAGGAGCTCGAATTTGCACTCTCCACACTCAGTTCCACAGATGTGTAGAGGAGT 60
Db 191265 CAGGTACGAGAGCTATATATCCCTGCTTTCACACTCAGCAACCTGCGAGATGAGAGT 191324

```

```

Qy 61 ATTCACTCCATTTTTCAGATGAGAGTGTAGGCCACAGAGAGCTAATCTGTCTGA 120
Db 191325 GACTATTTCCATCTTTAGATGAGAGATCAAAAGCTCAGAGAGGTTAAATGCTGCTAA 191384

```

```

Qy 121 GGCACACAGCTAGAAAGAGCC 143
Db 191385 GGCACACAGAGAGTGCAGGC 191407

```

```

RESULT 9
US-09-949-016-17033
; Sequence 17033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

```

FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 17033  
LENGTH: 194937  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-17033

Query Match 4.5%; Score 47; DB 4; Length 194937;  
Best Local Similarity 58.0%; Pred. No. 0.023;  
Matches 83; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 1 CAGGAGCTGAAATTGCACTCTCCACACTCAGTTCACAGATGTGTAGAGGCGAT 60  
Db 191265 CAGGTAGGAGGCTTATATCCCTGCTTCACACTCAGCAACCTGCGAGTAGAGT 191324  
Qy 61 ATTCACTCCATTTTTCAGATGAGAGTGGAGCCGAGAACTAGTAATCTGTCTGA 120  
Db 191325 GACTATTTCCATTCTTTAGATGAGATCAAGCTCAGAGAGGTAAATTAATCTGCTTA 191384  
Qy 121 GGCCACACAGCTAGAAAGAGCC 143  
Db 191385 GGCCACACAGCAGAGACTGCAAGC 191407

RESULT 10  
US-09-949-016-184953  
Sequence 184953, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 184953  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-184953

Query Match 4.5%; Score 46.8; DB 4; Length 601;  
Best Local Similarity 61.5%; Pred. No. 0.0023;  
Matches 75; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 22 CCTCCACACTCACTTCCACAGATGTGTAGAGGCGCATTTTCAATTTTTCAGAT 81  
Db 19 CGTGCTTTACACTCAGAGCAACCTGCGAGTAGAGTATTTCCATTTCTTTAGAT 78  
Qy 82 GAGAGTTGAGGCCGAGAACTAGTAATCTGTCTGAGGCCACAGCTAGAAAGCAG 141  
Db 79 GAGGATCAAGCTCAGAGAGGTAAATTAATCTGCTTAAGGCCACACAGAGAGACTGCA 138  
Qy 142 CC 143

Db 139 GC 140

RESULT 11  
US-09-949-016-185108  
Sequence 185108, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 185108  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-185108

Query Match 4.5%; Score 46.8; DB 4; Length 601;  
Best Local Similarity 61.5%; Pred. No. 0.0023;  
Matches 75; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 22 CCTCCACACTCACTTCCACAGATGTGTAGAGGCGCATTTTCAATTTTTCAGAT 81  
Db 19 CGTGCTTTACACTCAGAGCAACCTGCGAGTAGAGTATTTCCATTTCTTTAGAT 78  
Qy 82 GAGAGTTGAGGCCGAGAACTAGTAATCTGTCTGAGGCCACAGCTAGAAAGCAG 141  
Db 79 GAGGATCAAGCTCAGAGAGGTAAATTAATCTGCTTAAGGCCACACAGAGAGACTGCA 138

Qy 142 CC 143  
Db 139 GC 140

RESULT 12  
US-09-949-016-16222/C  
Sequence 16222, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 16222  
LENGTH: 43562  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(43562)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-16222

Query Match 4.5%; Score 46.8; DB 4; Length 43562;  
Best Local Similarity 63.2%; Pred. No. 0.014;  
Matches 72; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 38 CCACAGATGTTGAGGAGGATTCAGTCCATTTTTCAGATGAGAGTTGAGGCCA 97  
Db 38106 CACAACTCTATGATTAACAAATTTTTCATTTTACAGTCAAGATTGAGGACA 38047  
Qy 98 GAGAACGTAATATCTCTGAGGCCACACAGCTAGAAAGCAGGCCAG 151  
Db 38046 GAGATGTTAGTAATTTTACCTAAGGTACACAACTAATTAATTTGACAGGCCAG 37993

RESULT 13  
US-09-313-300-6  
Sequence 6, Application US/09313300  
Patent No. 6222027  
GENERAL INFORMATION:  
APPLICANT: Kaser, Matthew, R.  
APPLICANT: Lal, Preeti  
APPLICANT: Yue, Henry  
APPLICANT: Tang, Tom, Y.  
APPLICANT: Baughn, Mariah, R.  
APPLICANT: Azimzai, Yalda  
TITLE OF INVENTION: MOLECULES EXPRESSED IN HIPPOCAMPUS  
FILE REFERENCE: PB-0012 US  
CURRENT APPLICATION NUMBER: US/09/313,300  
CURRENT FILING DATE: 1999-05-17  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PERL Program  
SEQ ID NO 6  
LENGTH: 703  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY:  
OTHER INFORMATION: 1941247  
PUBLICATION INFORMATION:  
US-09-313-300-6

Query Match 4.5%; Score 46.2; DB 3; Length 703;  
Best Local Similarity 67.2%; Pred. No. 0.0036;  
Matches 80; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

Qy 30 CTCAGTCCACAGATGTTGAGGAGCATATTCAGTCCATTTTTCAGATGAG-AGT 88  
Db 510 CTCACACACCCCTGTGAGGTAGTCAATTTATTTATCCCATTTTACAGATGGGAAAC 569  
Qy 89 TGAGGCCAGAGACGTAATCTGTCTGAGGCCACACAGCTAGAAAGCAGCGCAGC 147  
Db 570 TGAGGCCAGAGACGTAATCTGTCTGAGGCCACACAGCGCAGCGCAGCGTCAAGC 628

RESULT 14  
US-09-949-016-14371  
Sequence 14371, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 14371  
LENGTH: 57837  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-14371

Query Match 4.5%; Score 46.2; DB 4; Length 57837;  
Best Local Similarity 58.3%; Pred. No. 0.023;  
Matches 81; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 31 TCAGTCCACAGATGTTGAGGAGCATATTCAGTCCATTTTTCAGATGAGAGTTG 90  
Db 20341 TGAATCTCAGAAATTAACCTTATGATGTCAGAGTCGTCATTTTACATTTAGAGCAG 20400  
Qy 91 AGGCCAGAGAGTAATATCTCTGAGGCCACACAGCTAGAAAGCAGGCCAGGCCA 150  
Db 20401 TGGCAGAGAAAGTAATTAATTTTGAAGTCAAGCCACACAGATCGTGGAGCCTGG 20460  
Qy 151 GCCGAACCCCTGCTGTGTG 169  
Db 20461 ACTGCTCTCAGCATGTG 20479

RESULT 15  
US-09-949-016-17601  
Sequence 17601, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17601  
LENGTH: 57839  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-17601

Query Match 4.5%; Score 46.2; DB 4; Length 57839;  
Best Local Similarity 58.3%; Pred. No. 0.023;  
Matches 81; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 31 TCAGTCCACAGATGTTGAGGAGCATATTCAGTCCATTTTTCAGATGAGAGTTG 90  
Db 20342 TGAATCTCAGAAATTAACCTTATGATGTCAGAGTCGTCATTTTACATTTAGAGCAG 20401  
Qy 91 AGGCCAGAGAGTAATATCTCTGAGGCCACACAGCTAGAAAGCAGGCCAGGCCA 150  
Db 20402 TGGCAGAGAAAGTAATTAATTTTGAAGTCAAGCCACACAGATCGTGGAGCCTGG 20461  
Qy 151 GCCGAACCCCTGCTGTGTG 169  
Db 20462 ACTGCTCTCAGCATGTG 20480

Search completed: June 29, 2005, 18:13:20  
Job time: 223 secs

**THIS PAGE BLANK (USPTO)**





D	b	61	ATTCTAGTCCCAATTTTTCAGATGAGGAGTTGAGGCCCAAGAAACGTAAATCTGTCTGA	120
Q	y	121	GGCCACACAGCTAAGAAAGCAGGCCAGGCCAGCCGAAACCTCTGTGTGTGCAGCCCCAGC	180
D	b	121	GGCCACACAGCTAAGAAAGCAGGCCAGGCCAGCCGAAACCTCTGTGTGTGCAGCCCCAGC	180
Q	y	181	CCAATTGCTCATTTGCGGGGCTCGGGAGCAAGAGGAAGCTGAGCAGCATGTGTTCCAGA	240
D	b	181	CCAATTGCTCATTTGCGGGGCTCGGGAGCAAGAGGAAGCTGAGCAGCATGTGTTCCAGA	240
Q	y	241	TGTTGGGAACTGAGAGAGCCCGCACAGGCCCTGTGCAAGGAAACCCGAGGGCTGTAGGC	300
D	b	241	TGTTGGGAACTGAGAGAGCCCGCACAGGCCCTGTGCAAGGAAACCCGAGGGCTGTAGGC	300
Q	y	301	CCCGTGGCACTGCATGTCTCAGGGCTGTGTGTCTGTGGAGCCACAGCCCTTCTGTACG	360
D	b	301	CCCGTGGCACTGCATGTCTCAGGGCTGTGTGTCTGTGGAGCCACAGCCCTTCTGTACG	360
Q	y	361	GCAGCAGAAATCTAGCCCGGAAAGGGTCAGGGAAATTCGTGAACATCTAGCAAGTCG	420
D	b	361	GCAGCAGAAATCTAGCCCGGAAAGGGTCAGGGAAATTCGTGAACATCTAGCAAGTCG	420
Q	y	421	GGCTGGGGTGTGGCCAAATTAGACACAGATGTAGGGCCCTTGATCTCAGAAATTTGGCAG	480
D	b	421	GGCTGGGGTGTGGCCAAATTAGACACAGATGTAGGGCCCTTGATCTCAGAAATTTGGCAG	480
Q	y	481	CTCTTTTGGCCAGAGGGGGCCAGCGCTGTGTCCGGGCTGGGGATGTCAAGAAAGGTCACCT	540
D	b	481	CTCTTTTGGCCAGAGGGGGCCAGCGCTGTGTCCGGGCTGGGGATGTCAAGAAAGGTCACCT	540
Q	y	541	GGGGGTCTTCCACTACACCCCGCTGTGACACTGTAGACCCCAAGGGCTCGAGGGAGCC	600
D	b	541	GGGGGTCTTCCACTACACCCCGCTGTGACACTGTAGACCCCAAGGGCTCGAGGGAGCC	600
Q	y	601	AGCTGGAGCCCATAGGAGAGGGGCAGTTCTCTCCTGTAAAGGGATTGTCTGTAGCATAG	660
D	b	601	AGCTGGAGCCCATAGGAGAGGGGCAGTTCTCTCCTGTAAAGGGATTGTCTGTAGCATAG	660
Q	y	661	GGAACAGACAAAGGCCAGGGGGAGCTAACCCGAGATCCAGCCCGGCTCACTCCGATG	720
D	b	661	GGAACAGACAAAGGCCAGGGGGAGCTAACCCGAGATCCAGCCCGGCTCACTCCGATG	720
Q	y	721	GCTACAGGCMAATATCTTAACCTCTCTGTAGACCTCTTGCCCAAGCTTACAGAGGTCATG	780
D	b	721	GCTACAGGCMAATATCTTAACCTCTCTGTAGACCTCTTGCCCAAGCTTACAGAGGTCATG	780
Q	y	781	AGGGGGTGAAGAAAGCCACGACGATGAAAGCTTTTAAACATTCTCGGGTGAAGCAGC	840
D	b	781	AGGGGGTGAAGAAAGCCACGACGATGAAAGCTTTTAAACATTCTCGGGTGAAGCAGC	840
Q	y	841	CCCTTCCCAATGCTGTGTCTGACTGCACGTGTGTGTGTAGGGGGTCCCAAGGGGCTC	900
D	b	841	CCCTTCCCAATGCTGTGTGTCTGACTGCACGTGTGTGTGTAGGGGGTCCCAAGGGGCTC	900
Q	y	901	AGTGTGGGCTGAGGCTGTGAACTAGGACAGGGGCTCTAGGAAAGAGCCTCTCTCC	960
D	b	901	AGTGTGGGCTGAGGCTGTGAACTAGGACAGGGGCTCTAGGAAAGAGCCTCTCTCC	960
Q	y	961	TGCCCACTGGGCATAGGCTCTGGAGCTGACAGCATGTGATCTCACTGATGCACTGG	1020
D	b	961	TGCCCACTGGGCATAGGCTCTGGAGCTGACAGCATGTGATCTCACTGATGCACTGG	1020
Q	y	1021	CCCTTCCCGCCAGGGCAG	1038
D	b	1021	CCCTTCCCGCCAGGGCAG	1038

```

RESULT 2
US-10-617-443B-1
; Sequence 1, Application US/10617443B
; Publication No. US20050019777A1
; GENERAL INFORMATION:

```

```

?   APPLICANT: Darrow, Andrew L
?   APPLICANT: Qi, Jian-shen
?   APPLICANT: Chen, Callin
?   APPLICANT: Andrade-Gordon, Patricia
?   TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
?   FILE REFERENCE: ORT-1644
?   CURRENT APPLICATION NUMBER: US/10/617,443B
?   CURRENT FILING DATE: 2003-07-02
?   NUMBER OF SEQ. ID NOS: 8
?   SOFTWARE: PatentIn Ver. 2.1
?   SEQ. ID NO 1
?   LENGTH: 3006
?   TYPE: DNA
?   ORGANISM: Homo sapiens
?   ORGANISM: Homo sapiens
US-10-617-443B-1

Query Match          100.0%;   Score 1038;   DB 21;   Length 3006;
Best Local Similarity 100.0%;
Matches 1038;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0.

```

Qy	1	CAGGGA	CTGAA	GGTTG	GAGT	CTC	CA	CA	CT	CA	AT	TC	CC	A	CA	GT	GT	GT	GT	AG	A	GG	CA	T	60												
Db	1	CAGGGA	CTGAA	GGTTG	GA	GT	CT	CT	CA	CA	CT	CA	AT	TC	CC	A	CA	GT	GT	GT	AG	A	GG	CA	T	60											
Qy	61	ATTGAG	TCCAT	TTTTT	TC	AAG	TGAG	GA	GT	TG	AG	GG	CC	CA	GAA	AC	GT	AA	TAT	CT	T	CT	GA		120												
Db	61	ATTGAG	TCCAT	TTTTT	TC	AAG	TGAG	GA	GT	TG	AG	GG	CC	CA	GAA	AC	GT	AA	TAT	CT	T	CT	GA		120												
Qy	121	GGCC	CA	CAG	CT	GA	AA	GA	CC	CA	AG	CC	CA	AG	CC	CA	AG	CC	CA	AG	CC	CA	AG		180												
Db	121	GGCC	CA	CAG	CT	GA	AA	GA	CC	CA	AG	CC	CA	AG	CC	CA	AG	CC	CA	AG	CC	CA	AG		180												
Qy	181	CGAG	TG	CT	CA	TT	GG	GG	GG	CT	CG	GG	AG	CA	GA	CG	CA	GG	CT	GA	CA	GA	CT	GT	TC	GA	240										
Db	181	CGAG	TG	CT	CA	TT	GG	GG	GG	CT	CG	GG	AG	CA	GA	CG	CA	GG	CT	GA	CA	GA	CT	GT	TC	GA	240										
Qy	241	TG	GT	GG	AAA	CT	GA	GA	GA	AG	CC	CG	CA	AG	CC	CG	TC	GA	GG	AA	CC	CG	GA	GG	CT	GT	AG	300									
Db	241	TG	GT	GG	AAA	CT	GA	GA	GA	AG	CC	CG	CA	AG	CC	CG	TC	GA	GG	AA	CC	CG	GA	GG	CT	GT	AG	300									
Qy	301	CCC	GT	GG	CCA	CT	GA	AG	CC	CT	CA	GG	CC	TG	GT	GT	CT	GG	AG	CA	CA	AG	CC	CT	CA	CT	GT	AG	360								
Db	301	CCC	GT	GG	CCA	CT	GA	AG	CC	CT	CA	GG	CC	TG	GT	GT	CT	GG	AG	CA	CA	AG	CC	CT	CA	CT	GT	AG	360								
Qy	361	GG	AG	CA	GA	GA	AT	CT	GA	AG	CC	CG	GG	AAA	GG	GT	CT	GA	GG	AA	GG	CT	GA	GG	AA	GG	CT	GA	420								
Db	361	GG	AG	CA	GA	GA	AT	CT	GA	AG	CC	CG	GG	AAA	GG	GT	CT	GA	GG	AA	GG	CT	GA	GG	AA	GG	CT	GA	420								
Qy	421	GG	CT	GG	GG	GT	GT	GG	CC	AA	GT	T	GA	CA	CA	CA	GA	T	T	AG	GG	CC	CT	GT	GA	CT	CA	GAA	ATT	GG	CAG	480					
Db	421	GG	CT	GG	GG	GT	GT	GG	CC	AA	GT	T	GA	CA	CA	CA	GA	T	T	AG	GG	CC	CT	GT	GA	CT	CA	GAA	ATT	GG	CAG	480					
Qy	481	CT	CT	TT	TT	GG	CC	CA	GA	GG	GG	CC	CA	AG	CT	GT	GT	CT	GG	GG	CC	CT	GG	GT	AG	CT	CA	GA	AA	GG	GT	CA	CT	540			
Db	481	CT	CT	TT	TT	GG	CC	CA	GA	GG	GG	CC	CA	AG	CT	GT	GT	CT	GG	GG	CC	CT	GG	GT	AG	CT	CA	GA	AA	GG	GT	CA	CT	540			
Qy	541	GG	GG	GG	CT	TT	CA	CT	CA	CA	CC	CG	CT	GA	CA	CT	GA	CA	CT	GA	CA	CT	GA	CA	CT	GA	CA	CT	GA	CA	CT	GA	CA	CT	600		
Db	541	GG	GG	GG	CT	TT	CA	CT	CA	CA	CC	CG	CT	GA	CA	CT	GA	CA	CT	GA	CA	CT	GA	CA	CT	GA	CA	CT	GA	CA	CT	GA	CA	CT	600		
Qy	601	AG	CT	GA	AG	CC	CA	TG	AG	AG	GG	CC	CA	AG	GG	CC	CA	AG	GG	CC	CA	AG	GG	CC	CA	AG	GG	CC	CA	AG	GG	CC	CA	AG	GG	CC	660
Db	601	AG	CT	GA	AG	CC																															

```

RESULT 4
US-09-764-891-5712
: Sequence, 5712, Application US/09764891
: Publication No. US20030077808A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC006
: CURRENT APPLICATION NUMBER: US/09/764,891
: CURRENT FILING DATE: 2001-01-17

```

```

RESULT 6
US-10-074-024-572/c
Sequence 572, Application US/10074024
Publication No. US20030232975A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC001C1
CURRENT APPLICATION NUMBER: US/10/074,024
PRIORITY FILING DATE: 2002-02-14
Prior Application removed - See file Wrapper or Palm
NUMBER OF SEQ ID NOS: 879
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 572
LENGTH: 532
TYPE: DNA
ORGANISM: Homo sapiens
US-10-074-024-572

```

```
Query Match 4.8%; Score 49.6; DB 17; Length 532;
Best Local Similarity 65.2%; Pred. No. 0.00011;
Matches 73; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 67 TCCCATTTTCAGATGAGAGTTGAGGCCAGAGAACGTAATCTGTCTGAGGCCAC 126
DB 197 TGCCTTTTGAAGATGAAGAGCTGAGGCTCAGAGAAAGCAAGTATTCACCTGAGGCCAC 138

QY 127 ACAGCTAGAAGACAGCCAGCCAGCCGAAACCTGCTGTGTGACGCCCA 178
DB 137 GAGCTAGCAGAGAGGAGCCAGCAGCTCACTAAGTCTGCTCTGCCA 86

RESULT 7
US-10-027-632-32570
Sequence 32570, Application US/10027632
Publication No. US2002019837A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 32570
LENGTH: 710
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(710)
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-32570

Query Match 4.8%; Score 49.6; DB 13; Length 710;
Best Local Similarity 72.2%; Pred. No. 0.00011;
Matches 78; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

QY 41 CAGATGTGTAGAGAGGCGATATTCAGTCCATTTTTCAGATGAGA-GTTGAGGCCAGA 99
DB 191 CAGCCTTGAGAGGGGTGATTTTATTCATTTTACATGAGACAGAGGCTCAGA 250

QY 100 GACGTAAGTAATCTGTCTGAGGCCACACAGCTAGAGAACGAGCCAGGC 147
DB 251 GAGGTCAAGTATTTGCCAGGCCACACAGCTAGAGAACTGCCAGGC 298

RESULT 8
US-10-027-632-32570
Sequence 32570, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632

Query Match 4.8%; Score 49.6; DB 17; Length 710;
Best Local Similarity 65.2%; Pred. No. 0.00011;
Matches 73; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 67 TCCCATTTTCAGATGAGAGTTGAGGCCAGAGAACGTAATCTGTCTGAGGCCAC 126
DB 197 TGCCTTTTGAAGATGAAGAGCTGAGGCTCAGAGAAAGCAAGTATTCACCTGAGGCCAC 138

QY 127 ACAGCTAGAAGACAGCCAGCCAGCCGAAACCTGCTGTGTGACGCCCA 178
DB 137 GAGCTAGCAGAGAGGAGCCAGCAGCTCACTAAGTCTGCTCTGCCA 86

RESULT 9
US-10-074-024-71/C
Sequence 71, Application US/10074024
Publication No. US20030232975A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC001C1
CURRENT APPLICATION NUMBER: US/10/074,024
CURRENT FILING DATE: 2002-02-14
PRIOR APPLICATION removed - See file Wrapper or Palm
NUMBER OF SEQ ID NOS: 879
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 71
LENGTH: 414
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (86)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (374)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (388)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (391)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (411)
OTHER INFORMATION: n equals a,t,g, or c
```

; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720

Thu Jun 30 08:48:36 2005

us-10-617-443b-1\_copy\_1\_1038.rnpb

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226737
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226737

Query Match
  4.6%; Score 48; DB 13; Length 585;
Best Local Similarity 58.7%; Pred. No. 0.0034;
Matches 81; Conservative 1; Mismatches 56; Indels 0; Gaps 0;

QY 48 GGTAGAGGCGCATATTCAGTCCATTTTTCAGATGAGAGTTGAGGCCCGCAGAACTGTA 107
DB 493 GGAGGTAGGAACCTATTATCCCATTTTCTAGATGAGAACTGAGGCTCAGGAGGAAA 434
QY 108 GTAATCTGTCTGAGGCGCACAGCTAGAAAGCAGCCGCGCAACCCCTGTGTG 167
DB 433 GAAATTTACCATGTGTCAAAAGCTAGTAGTAGCTCAAAATCCAAACCCAGGTATG 374
QY 168 TGCAGCCCCCAGCCCACT 185
DB 373 CGTAACACCAAGCCAGT 356

RESULT 14
US-10-027-632-226738/c
; Sequence 226738, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226738
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226738

Query Match
  4.6%; Score 48; DB 13; Length 585;
Best Local Similarity 58.7%; Pred. No. 0.0034;
Matches 81; Conservative 1; Mismatches 56; Indels 0; Gaps 0;

QY 48 GGTAGAGGCGCATATTCAGTCCATTTTTCAGATGAGAGTTGAGGCCCGCAGAACTGTA 107
DB 493 GGAGGTAGGAACCTATTATCCCATTTTCTAGATGAGAACTGAGGCTCAGGAGGAAA 434
QY 108 GTAATCTGTCTGAGGCGCACAGCTAGAAAGCAGCCGCGCAACCCCTGTGTG 167
DB 433 GAAATTTACCATGTGTCAAAAGCTAGTAGTAGCTCAAAATCCAAACCCAGGTATG 374
QY 168 TGCAGCCCCCAGCCCACT 185
DB 373 CGTAACACCAAGCCAGT 356
```

```

RESULT 15
US-10-027-632-226739/c
; Sequence 226739, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226739
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226739

Query Match
  4.6%; Score 48; DB 13; Length 585;
Best Local Similarity 58.7%; Pred. No. 0.0034;
Matches 81; Conservative 1; Mismatches 56; Indels 0; Gaps 0;

QY 48 GGTAGAGGCGCATATTCAGTCCATTTTTCAGATGAGAGTTGAGGCCCGCAGAACTGTA 107
DB 493 GGAGGTAGGAACCTATTATCCCATTTTCTAGATGAGAACTGAGGCTCAGGAGGAAA 434
QY 108 GTAATCTGTCTGAGGCGCACAGCTAGAAAGCAGCCGCGCAACCCCTGTGTG 167
DB 433 GAAATTTACCATGTGTCAAAAGCTAGTAGTAGCTCAAAATCCAAACCCAGGTATG 374
QY 168 TGCAGCCCCCAGCCCACT 185
DB 373 CGTAACACCAAGCCAGT 356

Search completed: June 29, 2005, 18:26:05
Job time : 762 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2005, 15:10:02 ; Search time 4078 Seconds  
(without alignments)  
9688.747 Million cell updates/sec

Title: US-10-617-443B-1\_COPY\_1\_1038

Perfect score: 1038  
Sequence: 1 cagcgactcgaagttgacg.....ggccctcccgccagcgacg 1038

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
EST:  
1: gb\_ests1.\*  
2: gb\_ests2.\*  
3: gb\_hnc.\*  
4: gb\_ests3.\*  
5: gb\_ests4.\*  
6: gb\_ests5.\*  
7: gb\_ests6.\*  
8: gb\_gsa1.\*  
9: gb\_gsa2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	55.8	5.4	421	1	AL119813 DKFZP761H
C 2	55.8	5.4	458	1	AL119782 DKFZP761F
C 3	54.8	5.3	925	1	CNS0091P
C 4	51.6	5.0	547	8	AO808226 HS 4554 A
C 5	50.6	4.9	383	2	BE099553 UI-R-BJI-
C 6	50.6	4.9	449	9	CC773356 CH240_17D
C 7	50	4.8	536	9	CE070706
C 8	49.6	4.8	689	9	AG113318 Pan tregl
C 9	48.6	4.7	790	6	CC552420 CH240_438
C 10	48	4.6	667	6	CB321964 UI-CF-FNO
C 11	46.8	4.5	454	8	AO415695
C 12	46.8	4.5	672	8	AO237490 RPII1-1-1
C 13	46.8	4.5	925	2	CNS0091P
C 14	46.6	4.5	150	2	BE145776
C 15	46.6	4.5	422	4	BI028582 IIS-HT020
C 16	46.6	4.5	510	8	AO479326 RPII-11-2
C 17	46.4	4.5	700	9	CE466727 ligR-g88-
C 18	46.2	4.5	643	7	CN070055 857546 MA
C 19	45.6	4.4	586	2	AM658542
C 20	45.4	4.4	256	1	AA256443 EST11220
C 21	45.2	4.4	662	9	AG051198 Pan tregl
C 22	45	4.3	280	1	AI702247 w604C04.x
C 23	45	4.3	462	1	AI884801 w184h03.x
C 24	44.8	4.3	506	8	AQ777662 HS_2252_B

25	44.6	4.3	460	2	BF921020
C 26	44.6	4.3	578	5	BO364741
C 27	44.6	4.3	575	5	BE109848
C 28	44.6	4.3	611	4	BG378055
C 29	44.6	4.3	668	1	AV221826
C 30	44.6	4.3	733	9	AG030822
C 31	44.6	4.3	762	6	CD777006
C 32	44.6	4.3	763	6	CD774640
C 33	44.4	4.3	253	1	AI560342
C 34	44.4	4.3	591	1	AU145409
C 35	44.4	4.3	750	9	CC531733
C 36	44.2	4.3	480	9	CE347071
C 37	44.2	4.3	910	9	CNS0060N
C 38	44.2	4.2	610	8	AO374046
C 39	44	4.2	717	9	CE168278
C 40	43.8	4.2	261	2	AM270824
C 41	43.8	4.2	742	5	BP159679
C 42	43.8	4.2	769	6	CD104579
C 43	43.8	4.2	836	9	CC485946
C 44	43.6	4.2	438	8	AO010437
C 45	43.6	4.2	504	8	B94987

## ALIGNMENTS

RESULT 1  
AL119813/c 421 bp mRNA linear EST 04-SEP-2003  
LOCUS DKFZP761H1124.1 761 (synonym: hamy2) Homo sapiens cDNA clone  
DEFINITION DKFZP761H1124.5, mRNA sequence.  
ACCESSION AL119813  
VERSION AL119813.1 GI:5925712  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Bloeker,H., Boecker,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and Wiemann,S.  
TITLE EST (Bloeker, et al.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: MIPS

MFPS  
Ingolstaedter Landstr.1, D-85764 Neuberg, Germany  
This is the 5' sequence of the clone insert  
clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
No 5' sequence available.  
This clone (DKFZP761H124) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de.  
Location/Qualifiers

FEATURES  
source  
1..421  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZP761H1124"  
/issue\_type="amalgam"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_idb="761 (synonym: hamy2)"  
/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"

## ORIGIN

Query Match 5.4%; Score 55.8; DB 1; Length 421;  
Best Local Similarity 70.1%; Pred. No. 0.0027;  
Matches 89; Conservative 0; Mismatches 37; Indels 1; Gaps 1;



QY	67	TCCCATTTTTGAGTGAAGATTGAGGCCGAGAAAGTAAGTATCTGTCTGAGCCAC	126
Db	279	TCCCATTTTTAGGTGTAAACTGAGGCCGAGAGCTTTATGTGACTTGTCTCAAGGTAC	220
QY	127	ACAGCTAAGAAAGCAGCCAGGCC-CAGCCGAACCCCTGTGTGTGCAGGCCCCAGCCAGT	185
Db	219	ACAGCTCAAAAGCAGCAAAAGCCGAAGCTGAAACCAGGTCGTCTCAAGCTCCCTGGCCAGG	160
QY	186	TGCTCAT	192
Db	159	CCATCTT	153

RESULT 2	AL119782/c	458 bp	mRNA	linear	EST 04-SEP-2003
LOCUS	AL119782/c				
DEFINITION	DKRZP61FI124.r1.761 (synonym: hamy2) Homo sapiens cDNA clone				
ACCESSION	AL119782				
VERSION	AL119782.1	GI:5925661			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 458) Bloecher,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and Wiemann,S				
TITLE	EST (Bloecher, et al.)				
JOURNAL	Unpublished (1995)				
COMMENT	Contact: MIPS				

Ingolstaedter Landstr.1, D-85764 Neuberg, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ), Email: a.wiemann@dkfz-heidelberg.de;  
sequenced by GBF (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the CGNA sequencing consortium of the  
German Genome Project.  
No BL sequence available.  
This clone (DKFZp611f124) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubergweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

```

FEATURES
source
Location/Qualifiers
1. .458
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="DKFZp761F1124"
/rnause_type="amygdala"
/dev_stage="adult"
/lab_host="DH10B"
/clone_1lb="761 (synonym: hamy2)"
/note="vector: pspori1; Site_1: Nci; Site_2: SalI"
ORIGIN

```

Query Match	5.4%	Score 55.8;	DB 1;	length 458;
Best Local Similarity	70.1%	Pred No. 0.0027;		
Matches 89;	Conservative	0;	Mismatches 37;	Indels 1;
				Gaps 1;

Accession	Sequence	Position
Oy	67 TCCCAATTTTTCAGATGAGAGTTGAGGCCACAGAAAGTAACTGTTCTAAGGCAC	156
Db	219 TCCCAATTTTTCAGGTGTGAAATCGAGGCCACAGAGCTTTAAGTGA	220
Oy	127 ACAGGTGAAAGCAGCCAGGCC-C-CAGCCGAACTCTGTGTGTGACGCCCCACGCCAGT	185
Db	219 ACAGGTCAAAACAGCGAAGCCGAACTGAAACCCAGGTCTGTCCAGTCTCCGTGGCCAGG	160

QY	186	TGCTCAT	192
Db	159	CCATCTT	153

RESULT 3	
CNS0091P/c	
LOCUS	CNS0091P 925 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TERT3 end of BAC #
	BAR18D16 of R601-98 library from Drosophila melanogaster (fruit
	fly), genomic survey sequence.
ACCESSION	AL053013
VERSION	AL053013.1
KEYWORDS	GI:4934461
SOURCE	GSS.
ORGANISM	Drosophila melanogaster (fruit fly)
	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
	Epitheroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 925)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

COMMENT

Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oosagawa and Aaron Mammone in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

FEATURES
source
location/Qualifiers
1..925
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_1b="RPC1-98"
/note="end : TET3"

```

Query Match	5.3%;	Score 54.8;	DB 9;	Length 925;
Best Local Similarity	16.5%;	Pred. No. 0 0057;		
Matches 60;	Conservative 161;	Mismatches 142;	Indels 1;	Gaps 1;

[illegible]

608 GCCCATGAGGAGAGGCCAGTTCCTCTCTGTAAAGGTATTGCTGTAGCATGAGGACAG 667

[illegible]

Qy 788 TGAGGAAACCCAGCACTGTGAAGCCTTTTAAACATTCCTGGGGTGAGCGAGCCCCCTTC 847  
::: ::::  
Db 615 VSASSGMSSTVSSSGGRSGSGGGGGTGVGSSSSSSAGSSSAGSSSVCSGSSGCRCRCS 556



QY 848 CAA 851  
Db 555 SAA 552

RESULT 4  
A0808226  
LOCUS  
DEFINITION  
A0808226 547 bp DNA linear GSS 10-AUG-1999  
HS 4554 A2 H04 T7A CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=4554 Col=8 Row=O, genomic survey  
sequence.

ACCESSION  
A0808226  
VERSION  
A0808226.1 GI:5727468  
KEYWORDS  
GSS.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens

REFERENCE  
AUTHORS  
Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,  
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and  
Hood L.  
Sequence-tagged connectors: A sequence approach to mapping and  
examining the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

JOURNAL  
MEDLINE  
PUBMED  
99380589  
10449764  
Contact: Mahairas G.G., Wallace J.C., Hood L.  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@reagen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 4554 row: 0 column: 8  
Seq primer: T7  
Clase: BAC ends  
High quality sequence stop: 547.  
Location/Qualifiers  
1. 547  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=4554 Col=8 Row=O"  
/sex="male"  
/clone\_id="CIT Approved Human Genomic Sperm Library D"  
/notes="Organ: sperm; Vector: pBelBAC11, BAC Clones in  
E-Coli DH10B"

FEATURES  
source  
1. 547  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=4554 Col=8 Row=O"  
/sex="male"  
/clone\_id="CIT Approved Human Genomic Sperm Library D"  
/notes="Organ: sperm; Vector: pBelBAC11, BAC Clones in  
E-Coli DH10B"

ORIGIN  
Query Match 5.0%; Score 51.6; DB 8; Length 547;  
Best Local Similarity 60.9%; Pred. No. 0.034;  
Matches 84; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 48 GGTGAGGCGCATATTCAGTCCATTTTTCAGATGAGAGTGGCCCGAGAACTTAA 107  
Db 225 GGAAGTGAAGCACTATTCCTCCATTTCTAGATGAAGAACTGAGGCTCAGGAGAGAAA 284  
QY 108 GTATCTGTCTGAGCCACAGCTAGAAAGCCGCGCCGCAACCCCTGTGTG 167  
Db 285 GAAATTTTACCACTGTCACAAAGCTAGTAGTAGTGAATTCACAAACCCAGTATG 344  
QY 168 TGCAGCCCGCAGCCACT 185  
Db 345 CGTAAACCAAGCCACT 362

RESULT 5  
BE099553 383 bp mRNA linear EST 13-JUN-2000  
LOCUS  
DEFINITION  
UI-R-BJ1-actj-f-11-0-UI.61. UI-R-BJ1 Rattus norvegicus cDNA clone

ACCESSION  
BE099553  
VERSION  
BE099553.1 GI:8491431  
KEYWORDS  
EST.  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus

REFERENCE  
AUTHORS  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)

JOURNAL  
MEDLINE  
PUBMED  
97044477  
8889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBR, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: Bento-soares@uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
M.B. Soares Lab Clone distribution: clones will be available  
through Research Genetics (www.reagen.com)  
Seq primer: M13 Forward  
POLY-A=yes.

FEATURES  
source  
1. 383  
Location/Qualifiers  
1. 383  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BJ1-actj-f-11-0-UI"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_id="UI-R-BJ1"  
/notes="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ1  
library is a subcloned library derived from the following  
tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV  
canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,  
AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.  
For a detailed description of the library from which this  
clone was derived, please visit our web site at  
ratseq.eng.uiowa.edu. The subtraction has been previously  
described in (Bonaldo, Lennon and Soares, Genome Research  
6:791-806, 1996)  
TAG TISSUE=ventricle at 16.5 dpc  
TAG\_LIB=UI-R-BJ1  
TAG\_SEQ=GTTCG"

ORIGIN  
Query Match 4.9%; Score 50.6; DB 2; Length 383;  
Best Local Similarity 70.1%; Pred. No. 0.058;  
Matches 68; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 51 AGAGGCGCATATTCAGTCCATTTTTCAGATGAGAGTGGCCCGAGAACTTAA 110  
Db 23 AGGTGTCATTTTATTCCTCCATTTCTAGTGAAGAACTGAGGCTCAGAGAACTG 82  
QY 111 ATCTGTGAGCGCCACAGCTAGAAAGCCGCGCCGCAACCCCTGTGTG 147  
Db 83 AGTGCACAGCCACAGCTAGAAAGTTCAGGCGC 119

RESULT 6  
CC773356

LOCUS CC773356 449 bp DNA linear GSS 27-JUN-2003  
 DEFINITION CH240\_17D11.TV CHORI-240 Bos taurus genomic clone CH240\_17D11,  
 genomic survey sequence.  
 ACCESSION CC773356  
 VERSION CC773356  
 KEYWORDS GI:32325647  
 SOURCE GSS.  
 ORGANISM Bos taurus (cow)  
 Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.  
 1 (bases 1 to 449)  
 Larkin,D.M., Evers-van der Wind,A., Rebeiz,M., Schweitzer,P.,  
 Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L.,  
 Womack,J.E., de Jong,P.V. and Lewin,H.A.  
 Bovine BAC end sequences from CHORI-240 library  
 Unpublished (2003)  
 CONTACT: Harris Lewin  
 Department of Animal Sciences  
 University of Illinois at Urbana Champaign  
 1201 W. Gregory Dr., Urbana, IL 61801, USA  
 Tel: 217 333 5998  
 Fax: 217 244 5617  
 Email: h-lewin@uiuc.edu  
 Clones are derived from the bovine BAC library CHORI-240  
 (<http://www.chori.org/bacpac/bovine240.htm>). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
 Clones may be purchased from BACPAC Resources  
 ([http://www.chori.org/bacpac/ordering\\_information.htm](http://www.chori.org/bacpac/ordering_information.htm)). This work  
 was undertaken as part of the International Bovine BAC Mapping  
 Consortium (IBMC) by the University of Illinois at Urbana  
 Champaign, USA with funds provided by grant No. AG202-34480-11828  
 from USDA-CSREBS and AG99-35205-8534 from USDA/NRI (Livestock  
 Genome Sequencing Initiative)  
 Plate: 17 row: D column: 11  
 Seq primer: T7  
 Class: BAC end.

FEATURES  
 source Location/Qualifiers  
 1..449  
 /organism="Bos taurus"  
 /mol\_type="genomic DNA"  
 /strain="Breed: Hereford"  
 /db\_xref="taxon:9913"  
 /clone="CH240\_17D11"  
 /sex="Male"  
 /cell\_type="Blood"  
 /clone\_lib="CHORI-240"  
 /note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;  
 Hereford bull LI Domino 99375; CHORI-240 Bovine BAC  
 library (Male) produced by Pieter de Jong"

ORIGIN  
 Query Match 4.9%; Score 50.6; DB 9; Length 449;  
 Best Local Similarity 60.6%; Pred. No. 0.06;  
 Matches 83; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 33 AGTTCCCAAGATGTGTAGAGGAGCATATTCAGTCCCTTTTCAGAGAGAGTTGAG 92  
 DB 81 AACTTGACACCGAAGCTGTAGAAACGTATGTAGCTTTTACGATGAGAACTGAG 140  
 QY 93 GCCAGAGAACGTAACTATCTGTGTAGAGCCACACAGCTAGAAAGCAGCCAGCCAC 152  
 DB 141 GCTCAGAGAGTTGAGTCACTTGTAGCTGAGGTCAACAGCTAGCACTGCTGGGG 200  
 QY 153 CGAACCCCTGTGTGTG 169  
 DB 201 AGAGAGCCGAGTCGCTG 217

RESULT 7  
 CE070706 536 bp DNA linear GSS 24-SEP-2003  
 LOCUS CE070706  
 DEFINITION tigr-gss-dog-17000323019295 Dog Library Canis familiaris genomic.

ACCESSION CE070706  
 VERSION CE070706.1 GI:35132048  
 KEYWORDS GSS.  
 SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 1 (bases 1 to 536)  
 Kirkness,E.F., Baltn,V., Halpern,A.L., Levy,S., Remington,K.,  
 Rutsch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
 Venter,J.C.  
 The dog genome: survey sequencing and comparative analysis  
 Science 301 (5641), 1898-1903 (2003)  
 22875432  
 14512627  
 CONTACT: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-838-0200  
 Fax: 301-838-0208  
 Email: ekirkness@tigr.org  
 Class: shotgun.

FEATURES  
 source Location/Qualifiers  
 1..536  
 /organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"  
 /db\_xref="taxon:9615"  
 /clone\_lib="Dog Library"  
 /note="Site 1: BstXI; Libraries were prepared from  
 peripheral blood"

ORIGIN  
 Query Match 4.8%; Score 50; DB 9; Length 536;  
 Best Local Similarity 60.1%; Pred. No. 0.089;  
 Matches 83; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 49 GTAGAGGCAATATTCAGTCCCATTTTCAGATGAGAGTTGAGCCCAAGAACTAG 108  
 DB 36 GAGAGAGATATTTTGACCTGATCTTCAAAAGAGAACTGAGACTCAGAGACCTTAG 95  
 QY 109 TAATGTGTGTAGAGCCACACAGCTAGAAAGCAGCCAGCCAGCCAGCCCTGTGTGT 168  
 DB 96 TAGCTAGCCCAAGGTCAACAGCTGTAGTGTGAGCCAGCTGAGATTCACTTGAATT 155  
 QY 169 GCAAGCCCAAGCCCAAGTT 186  
 DB 156 GAGGCCCAATCCGCTGCT 173

RESULT 8  
 AG113318 689 bp DNA linear GSS 03-NOV-2001  
 LOCUS AG113318  
 DEFINITION Pan troglodytes DNA, clone: PTB-119015.R, genomic survey sequence.  
 ACCESSION AG113318  
 VERSION AG113318.1 GI:16733837  
 KEYWORDS GSS.  
 SOURCE Pan troglodytes (chimpanzee)  
 ORGANISM Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
 1  
 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
 Totoki,Y., Watanabe,H. and Sakaki,Y.  
 BAC end sequences of library PTB  
 Unpublished  
 2 (bases 1 to 689)  
 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
 Totoki,Y., Watanabe,H. and Sakaki,Y.  
 Direct Submission  
 Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: [chimbases@sc.riken.go.jp](mailto:chimbases@sc.riken.go.jp), URL: <http://hnp.gsc.riken.go.jp/>,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PBH This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

## PRIMERS

Sequencing: M13Rev

## LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. 689

/organism="Pan troglodytes"

/mol\_type="genomic DNA"

/db\_xref="taxon:9598"

/clone="PTB-119015.R"

/sex="male"

/cell\_type="Tymphoblast"

/clone\_lib="PTB Chimpanzee Male BAC Library"

## ORIGIN

Query Match 4.8%; Score 49.6; DB 9; Length 689;  
Best Local Similarity 72.2%; Pred. No. 0.12; Mismatches 29; Indels 1; Gaps 1;  
Matches 78; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

QY 41 CAGATGTGTAGAGGCGCATTTTCATTTTTCATGATGAGA-GTTAGAGCCAGA 99

DB 185 CAGCTTGTAGAGGCGCATTTTTCATTTTTCATTTTACAGATGAGACAGAGGCTCAGA 244

QY 100 GAAGCTAGTATATCTCTGAGGCGCACAGCTGAGAAAGCCAGCCAGGC 147

DB 245 GAGGTCAATGATTTGCCAGGCGCACAGCTGAGAAAGTCCAGGC 292

## RESULT 9

CCS52420/c

LOCUS CH240\_438G3\_TARBA13P2 CHORI-240 Bos taurus genomic clone

DEFINITION CH240\_438G3, genomic survey sequence.

ACCESSION CCS52420

VERSION CCS52420.1 GI:31870704

KEYWORDS GSS.

SOURCE Bos taurus (cow)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.

1 (bases 1 to 790)

Holt, R., Stott, J., Yang, G., Barber, S., Smailus, D., Prabhu, A.-L.,  
Teal, M., Cloutier, A., Lee, D., Gilm, N., Olson, T., Mayo, M.,  
Butterfield, Y., Kirkpatrick, R., Liu, J., Guin, R., Chan, A., Chiu, R.,  
Mathewson, C., Wye, N., Masson, A., Brown-John, M., Jones, S.,  
Schein, J., Marra, M., de Jong, P., Keefe, J. W. and Kappe, S. M.

Bovine BAC End Sequences from Library CHORI-240, PLATES 399 to 478  
Unpublished (2003)

Other GSSs: CH240\_438G3.T7

Contact: Rob Holt

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

The British Columbia Cancer Agency Genome Science Centre  
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4S6  
Tel: 604-877-6085  
Fax: 604-877-6276  
Email: [rholt@bccsc.ca](mailto:rholt@bccsc.ca)

Clones are derived from the bovine BAC library CHORI-240  
(<http://www.chori.org/bacpac/bovine240.htm>). For BAC library  
availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).  
Clones may be purchased from BACPAC Resources  
(<http://www.chori.org/bacpac/ordering/information.htm>). This work  
was undertaken as part of the International Bovine BAC Mapping  
Consortium (IBMC) by CSIRO Livestock Industries, Australia and the

British Columbia Genome Sciences Centre, Canada.  
Plate: 438 row: G column: 3  
Seq primer: SP6  
Class: BAC ends

## FEATURES

source

Location/Qualifiers

1. 790

/organism="Bos taurus"

/mol\_type="genomic DNA"

/strain="Breed: Hereford"

/db\_xref="taxon:9913"

/clone="CH240\_438G3"

/sex="Male"

/cell\_type="Blood"

/clone\_lib="CHORI-240"

/note="Vector: pTARBA1.3; Site 1: MboI, Site 2: MboI;  
Hereford bull LI Domino 99175; CHORI-240 Bovine BAC  
library (Male) produced by Pieter de Jong"

## ORIGIN

Query Match 4.7%; Score 48.6; DB 9; Length 790;  
Best Local Similarity 55.7%; Pred. No. 0.22; Mismatches 74; Indels 0; Gaps 0;  
Matches 93; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 48 GGTAGAGGCGCATATTCATGTCCTTTTTCATGATGAGAGTTCAGAGCCAGACGTAA 107

DB 646 GCGAGTAGGATTTCTTATCCCATTTTTCAGATGATGATGAACTCAGAGGCTTA 587

QY 108 GTAATCTGTGAGGCCACACAGCTGAGAAAGCCAGCCAGCCAGCCAGCTGTGTG 167

DB 586 GTCACTTGCCCATATGACACAGCTGTATGCAATTCCTGCTCAAAACCAAGCTGT 527

QY 168 TGCAGCCCGCAGCCCACTTCTCATTTGCGGGGCTCGGAGCCAGAG 214

DB 526 TACTGACTCCCATGATGAGGTTTTCATATGATGCTTACAG 480

## RESULT 10

CB321964

LOCUS CB321964 667 bp mRNA linear EST 05-MAR-2003

DEFINITION UI-CF-FNO-aem-m-06-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone

UI-CF-FNO-aem-m-06-0-UI 3', mRNA sequence.

ACCESSION CB321964

VERSION CB321964.1 GI:28856622

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 667)

Normalization and subtraction: two approaches to facilitate gene  
discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

COMMENT

McCray, PB

McCray Lab

University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171

Email: [paul-mccray@uiowa.edu](mailto:paul-mccray@uiowa.edu)

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics ([www.resgen.com](http://www.resgen.com)) or from Open Biosystems

([www.openbiosystems.com](http://www.openbiosystems.com)).

The following repetitive elements were found in this cDNA

sequence: 174-405, >MIR5INE/MIR (matched complement)

Seq primer: M13 FORWARD

POLYA=Yes.

FEATURES  
Source

Location/Qualifiers  
1..667  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-FNO-aem-m-06-0-UI"  
/issue\_type="Human Lung Epithelial cells"  
/lab\_host="PHIOB (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-FNO"  
/note="Organ: Lung; Vector: pT73-Pac (pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-FNO is a subcloned cDNA library derived from two normalized Human lung epithelial cell libraries (EN) and D11. The library was subcloned according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. For additional information, contact: benito-soares@uiowa.edu  
TAG TISSUE=Lung Epithelial cells Tissue nos 359-368  
TAG\_LIB=UI-CF-FNO  
TAG\_SEQ=GGCTGTAGGC"

## ORIGIN

Query Match 4.6%; Score 48; DB 6; Length 667;  
Best Local Similarity 59.6%; Pred. No. 0.3; 55; Indels 0; Gaps 0;  
Matches 81; Conservative 0; Mismatches 0;

QY 51 AGNAGGCGCATTTTCAGTCCATTTTCAGATGAGAGTTCAGCCGAGAACTGTA 110  
DB 264 AGATGGAGAAATCAGCCCACTTTGAGATGATCTTGAAAGCTCAGAGGGTACAGA 323  
QY 111 ATCTGTCTGAGGACACAGCTAGAAAGCAGCCGAGCCGAAACCTGTGTGTGC 170  
DB 324 AGTTGTCTGAGACTCGCAGGGAGAGATGCTGAGCCATTTTCAMACAGCAGTCC 383  
QY 171 AGCCCCAGCCCACTT 186  
DB 384 TGCATACAGTCCGTT 399

RESULT 11  
LOCUS AQ415695 454 bp DNA linear GSS 23-MAR-1999  
DEFINITION RPCI-11-196N2.TV RPCI-11 Homo sapiens genomic clone RPCI-11-196N2,  
genomic survey sequence.  
ACCESSION AQ415695  
VERSION AQ415695.1 GI:4474664  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 454)  
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and  
Venter,J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
JOURNAL Unpublished (1997)  
COMMENT Other GSSs: RPCI-11-196N2.TJ  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buflalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buflalo.edu/ordering) or from  
Research Genet cs (info@resgen.com). BAC end search page:  
http://www.tigr.org/cdb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: T7  
Class: BAC ends.

FEATURES  
Source

Location/Qualifiers  
1..454  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7575193"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-196N2"  
/sex="Male"  
/cell\_type="lymphocytes"  
/clone\_lib="RPCI-11"  
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
RPCI11 Human Male BAC library"

## ORIGIN

Query Match 4.5%; Score 46.8; DB 8; Length 454;  
Best Local Similarity 58.7%; Pred. No. 0.57; 57; Indels 0; Gaps 0;  
Matches 81; Conservative 0; Mismatches 0;

QY 48 GGTAGAGGCGATTTTCAGTCCATTTTCAGATGAGAGTTCAGCCGAGAACTGTA 107  
DB 165 GAGATGAGAACTTTATCCCATTTTCTGATGAGAACTGAGGCTCAGGAGAGAA 224  
QY 108 GTAATCTGTCTGAGGACACAGCTAGAAAGCAGCCGAGCCGAAACCTGTGTG 167  
DB 225 GAAATTTACCATGCTCAGAAAGCTAGTGAAGTGAAGTCAATCCAAACCAAGTATG 284  
QY 168 TGCAGCCCGCAGCCAGT 185  
DB 285 CGTTACACCAAGCCAGT 302

RESULT 12  
LOCUS AQ237490 672 bp DNA linear GSS 21-APR-1999  
DEFINITION RPCI11-69K22.TK RPCI-11 Homo sapiens genomic clone RPCI-11-69K22,  
genomic survey sequence.  
ACCESSION AQ237490  
VERSION AQ237490.1 GI:3669781  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 672)  
AUTHORS Berry,K., Rounalev,S.D., Zhao,S., Base,S., Linher,K., Golden,K.,  
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.  
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building  
JOURNAL Unpublished (1998)  
COMMENT Other\_GSSs: RPCI11-69K22.TJ  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mda@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buflalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buflalo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/cdb/hungen/bac\_end\_search/bac\_end\_search.html  
Seq primer: T7  
Class: BAC ends.

FEATURES  
Source

Location/Qualifiers  
1..672  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7526373"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-69K22"  
/sex="Male"  
/cell\_type="lymphocytes"

/none+Oxygen, head, neck; Vector: puc18, Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from OHSR5 PCR (U.S. Letters Patent application No. 196, 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under

ORIGIN low stringency conditions."

Query Match 4.5%; Score 46.6; DB 2; Length 150;  
Best Local Similarity 71.8%; Pred. No. 0.51;  
Matches 61; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
QY 51 AGGAGGGCATATTCAGTCCCATTTTTCAGATGAGAGTTGAGCCCGAGAGAACTGTAAGT 110  
DB 49 AGGTAGGCACTATTAGCCCATTTTACAGATGAGAGAAAGCTCAGAGAGTTTAACTA 108  
QY 111 ATCTGTGAGGCCACAGAGCTAGA 135  
DB 109 ATTGTCCAGGTCCACAGCAGAGA 133

RESULT 15  
BI028582/c 422 bp mRNA linear EST 14-JUN-2001  
LOCUS BI028582  
DEFINITION IL5-MT0208-210201-356-f06 MT0208 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BI028582  
VERSION BI028582.1 GI:14435212  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 422)  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.G.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
PUBMED 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=IL5&t=IL5-MT0208-  
210201-356-f06&t3=2001-02-21&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 320.  
Location/Qualifiers  
1..422  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="MT0208"  
/note="Organ: marrow; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORFESTS PCR (U.S. letters Patent application No.  
196,716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

QY 62 TTCAGTCCCATTTTTCAGATGAGAGTTGAGCCCGAGAGAACTGTAATCTGTCTGAG 121  
DB 206 TACAATTCATTTTTCAGATCAGAGAAACCGGCTCAGAGAACTTAACTTGTCCCAAG 147  
QY 122 GCCACACAGCTAGAAAG 138  
DB 146 GTCCACACAGCAGTAG 130  
Search completed: June 29, 2005, 18:09:38  
Job time : 4087 secs

ORIGIN  
Query Match 4.5%; Score 46.6; DB 4; Length 422;  
Best Local Similarity 75.3%; Pred. No. 0.64; 19; Indels 0; Gaps 0;  
Matches 58; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_p2n model

Run on: June 29, 2005, 19:59:10 ; Search time 202.461 Seconds

(without alignments)  
2153.984 Million cell updates/sec

Title: US-10-617-443B-2\_COPY\_1\_9  
Perfect score: 44  
Sequence: 1 MHLALPASA 9

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 segs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame\_p2n.model -DEV=xlh  
-Q=cgnt2\_USPRO.epool/US10617443/runat.29062005\_135931\_26061/app\_query.fasta\_1.718  
-DB=GenEmbl -QFMT=faetap -SUFFIX=rge -MIMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10617443 @CGN 1.1.3890 @runat.29062005\_135931\_26061 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_da.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pac.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	60597	9 AC113611	AC113611 Homo sapi
2	39	88.6	35546	3 AC074202	AC074202 Leishmani
3	39	88.6	110000	1 BX571965_13	Continuation (14 o
4	39	88.6	110000	1 CP000010_19	Continuation (20 o

C 5	39	88.6	117146	2 AC087796	AC087796 Leishmani
C 6	39	88.6	142401	8 AC118669	AC118669 Genomic s
C 7	39	88.6	150106	2 AC144913	AC144913 Mus muscu
C 8	39	88.6	160650	3 AC008188	AC008188 Drosophil
C 9	39	88.6	160885	2 AC018184	AC018184 Drosophil
C 10	39	88.6	167588	8 AF461424	AF461424 Oryza sat
C 11	39	88.6	196178	10 AC107764	AC107764 Mus muscu
C 12	39	88.6	218827	2 AC146955	AC146955 Otollemur
C 13	39	88.6	221049	2 AC137178	AC137178 Rattus no
C 14	39	88.6	221586	2 BX957332	BX957332 Datio rer
C 15	39	88.6	238103	2 AC099140	AC099140 Rattus no
C 16	39	88.6	246934	2 AC094675	AC094675 Rattus no
C 17	39	88.6	254152	2 AC134753	AC134753 Rattus no
C 18	39	88.6	306854	3 AE003797	AE003797 Drosophil
C 19	38	86.4	951	3 AY486427	AY486427 Trichinel
C 20	38	86.4	1963	10 HAMALIN3	M73991 Mesocricetu
C 21	38	86.4	38200	9 AC000046	AC000046 Homo sapi
C 22	38	86.4	60940	6 AX695473	AX695473 Sequence
C 23	38	86.4	119903	9 AC091866	AC091866 Homo sapi
C 24	38	86.4	120472	9 AC091834	AC091834 Homo sapi
C 25	38	86.4	143298	8 AC134240	AC134240 Oryza sat
C 26	38	86.4	146438	9 HSBAL9119	AL117329 Human DNA
C 27	38	86.4	152768	2 AC120512	AC120512 Rattus no
C 28	38	86.4	157928	10 AC112256	AC120512 Mus muscu
C 29	38	86.4	157944	9 AC092532	AC092532 Papio anu
C 30	38	86.4	159717	2 AC008368	AC008368 Trypanoso
C 31	38	86.4	160584	2 AC091655	AC091655 Trypanoso
C 32	38	86.4	162045	2 AC023488	AC023488 Trypanoso
C 33	38	86.4	170806	2 CR627483	CR627483 Datio rer
C 34	38	86.4	177312	5 BX005140	BX005140 Zebrafish
C 35	38	86.4	179599	2 AC112930	AC112930 Mus muscu
C 36	38	86.4	186315	2 AC110188	AC110188 Mus muscu
C 37	38	86.4	187939	5 AL928672	AL928672 Zebrafish
C 38	38	86.4	190201	2 AC129837	AC129837 Papio anu
C 39	38	86.4	195685	2 AC119570	AC119570 Papio ham
C 40	38	86.4	195680	2 AC116045	AC116045 Papio ham
C 41	38	86.4	196078	10 AL663091	AL663091 Mouse DNA
C 42	38	86.4	197796	2 AC129792	AC129792 Rattus no
C 43	38	86.4	202339	9 AC090966	AC090966 Papio anu
C 44	38	86.4	205928	9 CNS07ERU	AL583722 Human chr
C 45	38	86.4	209060	9 AC010733	AC010733 Homo sapi

## ALIGNMENTS

RESULT 1	AC113611	60597 bp	DNA	linear	PRI 10-MAR-2003
LOCUS	AC113611				
DEFINITION	Homo sapiens BAC clone Rpl1-421M20 from 4, complete sequence.				
ACCESSION	AC113611				
VERSION	AC113611.3	GI:28475761			
KEYWORDS	HTG:				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Isak,A. and Colton,M.				
TITLE	The sequence of Homo sapiens BAC clone Rpl1-421M20				
JOURNAL	Unpublished (2001)				
REFERENCE	3. (bases 1 to 60597)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-MAR-2002) Genome Sequencing Center, Washington				
REFERENCE	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	4. (bases 1 to 60597)				

AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (11-APR-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 5 (bases 1 to 60597)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 6 (bases 1 to 60597)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (10-MAR-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Feb 23, 2003 this sequence version replaced g1:20136967.  
COMMENT ----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0421M20  
-----  
NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.  
  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.  
  
MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see <http://genome.wustl.edu/gsc>  
  
SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male  
donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E.,  
Tateno,M., Caranese,J.J. and de Jong,P.J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
and coworkers at <http://www.chori.org>  
VECTOR: pBACe3.6  
  
NEIGHBORING SEQUENCE INFORMATION:  
This sequence is not the entire insert of the clone. This clone is  
overlapped by AC104650 and AC104825.  
  
Discrepant bases between AC104650, AC104825 and clone sequence.  
  
Data from AC104825 was used to finish this clone.  
  
FEATURES  
source  
1..60597  
    /organism="Homo sapiens"  
    /mol\_type="genomic DNA"  
    /db\_xref="taxon:9606"  
    /chromosome="4"  
    /map="4"  
    /clone="RP11-421M20"  
    /clone\_11b="RPCI-11"  
    539..903  
    repeat\_region  
    /rpc\_family="ERV1"  
    repeat\_region  
    /rpc\_family="ERV1"

repeat\_region 2797..2907  
    /rpc\_family="CAGAGA)n"  
repeat\_region 2908..3114  
    /rpc\_family="ERV1"  
repeat\_region 3122..3184  
    /rpc\_family="CAGAGA)n"  
repeat\_region 3256..3368  
    /rpc\_family="L1"  
repeat\_region 3500..3682  
    /rpc\_family="L1"  
repeat\_region 3713..3782  
    /rpc\_family="AT-rich"  
repeat\_region 5208..5410  
    /rpc\_family="MIR"  
repeat\_region 5451..5503  
    /rpc\_family="MIR"  
repeat\_region 5504..5677  
    /rpc\_family="MIR"  
repeat\_region 6645..6687  
    /rpc\_family="MIR"  
repeat\_region 6998..7067  
    /rpc\_family="MIR"  
repeat\_region 7741..8173  
    /rpc\_family="MER1\_type"  
repeat\_region 8530..8610  
    /rpc\_family="MIR"  
repeat\_region 8638..9514  
    /note="CPG island (%GC=76.7, o/e=0.80, #CPGs=114)"  
repeat\_region 9027..9048  
    /rpc\_family="GC-rich"  
repeat\_region 9063..9111  
    /rpc\_family="GC-rich"  
repeat\_region 9612..9739  
    /rpc\_family="MIR"  
repeat\_region 10414..10509  
    /rpc\_family="MER103"  
repeat\_region 10688..11034  
    /rpc\_family="MALR"  
repeat\_region 11781..11997  
    /rpc\_family="L2"  
repeat\_region 12212..12325  
    /rpc\_family="MIR"  
repeat\_region 13275..13625  
    /rpc\_family="L2"  
repeat\_region 13703..13769  
    /rpc\_family="CR1"  
repeat\_region 14926..15054  
    /rpc\_family="MIR"  
repeat\_region 15239..15281  
    /rpc\_family="T1p100"  
repeat\_region 15349..15515  
    /rpc\_family="MIR"  
repeat\_region 16000..16113  
    /rpc\_family="MIR"  
repeat\_region 16639..16917  
    /rpc\_family="L2"  
repeat\_region 17127..17228  
    /rpc\_family="T1p100"  
repeat\_region 17883..17982  
    /rpc\_family="MIR"  
repeat\_region 18108..18173  
    /rpc\_family="CT-rich"  
repeat\_region 19059..19127  
    /rpc\_family="L2"  
repeat\_region 20087..20234  
    /rpc\_family="MIR"  
repeat\_region 20576..20738  
    /rpc\_family="MIR"  
repeat\_region 22242..22380  
    /rpc\_family="(TGA)n"  
repeat\_region 23733..23804  
    /rpc\_family="L2"  
repeat\_region 23890..23987



```

repeat_region /rpt_family="MIR"
                24011..24440
                /rpt_family="MALR"
repeat_region 25254..25412
                /rpt_family="MIR"
repeat_region 26001..26150
                /rpt_family="MIR"
repeat_region 26473..26512
                /rpt_family="MIR"
repeat_region 26513..26894
                /rpt_family="MALR"
repeat_region 26895..27048
                /rpt_family="MIR"
repeat_region 27423..27724
                /rpt_family="MALR"
repeat_region 27725..28016
                /rpt_family="MALR"
repeat_region 28029..28096
                /rpt_family="MALR"
repeat_region 28099..28399
                /rpt_family="TTC"n"
repeat_region 28406..28597
                /rpt_family="MIR"
repeat_region 29613..29783
                /rpt_family="L1"
repeat_region 29988..30078
                /rpt_family="MIR"
repeat_region 30801..30893

```

Alignment Scores:  
 Pred. No.: 491  
 Score: 44.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 9  
 DB: 9

US-10-617-443B-2\_COPY\_1\_9 (1-9) x AC113611 (1-60597)

QY 1 MethiShleuAlaUpProAlaSerAla 9  
 DB 21559 ATGACCTGGCCCTTCCCGCAGCGCA 21585

1024

RESULT 2  
 AC074202 35546 bp DNA linear INV 28-JUN-2002  
 LOCUS Leishmania major chromosome 35 clone Ia946 strain Friedlin,  
 AC074202 complete sequence.  
 AC074202.5 GI:21622732  
 HTG.  
 Leishmania major  
 Leishmania major  
 Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae;  
 Leishmania.  
 1 (bases 1 to 35546)  
 Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,  
 McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.  
 Direct Submission  
 Submitted (18-JUL-2000) Seattle Biomedical Research Institution, 4  
 Nickerson Street, Seattle, WA 98109-1651, USA  
 2 (bases 1 to 35546)  
 Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,  
 McDonagh,P., Ivens,A., Nguyen,D., Munden,H., Stuart,K., Nelson,S.,  
 Marty,A., Seyler,A. and Mack,J.  
 Direct Submission  
 Submitted (09-JUN-2001) Seattle Biomedical Research Institution, 4  
 Nickerson Street, Seattle, WA 98109-1651, USA  
 3 (bases 1 to 35546)  
 Worthey,E.A., Sisk,E., Seyler,A., Nelson,S., Vogt,C., Robertson,L.,  
 Pentony,M., Ivens,A., Mack,J., Munden,H., Marty,A., Stuart,K. and  
 Myler,P.J.  
 Direct Submission

JOURNAL Submitted (21-JUN-2001) Seattle Biomedical Research Institution, 4  
 Nickerson Street, Seattle, WA 98109-1651, USA  
 REFERENCE 4 (bases 1 to 35546)  
 AUTHORS Worthey,E.A., Sisk,E., Seyler,A., Nelson,S., Vogt,C., Robertson,L.,  
 Pentony,M., Ivens,A., Mack,J., Munden,H., Marty,A., Stuart,K. and  
 Myler,P.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-NOV-2001) Seattle Biomedical Research Institution, 4  
 Nickerson Street, Seattle, WA 98109-1651, USA  
 REFERENCE 5 (bases 1 to 35546)  
 AUTHORS Worthey,E.A., Sisk,E., Seyler,A., Nelson,S., Vogt,C., Robertson,L.,  
 Pentony,M., Ivens,A., Mack,J., Munden,H., Marty,A., Stuart,K. and  
 Myler,P.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JUN-2002) Seattle Biomedical Research Institution, 4  
 Nickerson Street, Seattle, WA 98109-1651, USA  
 COMMENT On Jun 28, 2002 this sequence version replaced gi:16751888.  
 FEATURES  
 source  
 1..35546  
 /organism="Leishmania major"  
 /mol\_type="genomic DNA"  
 /strain="Friedlin"  
 /db\_xref="taxon:5664"  
 /chromosome="35"  
 /clone="I4946"

# ORIGIN

Alignment Scores:  
 Pred. No.: 3,73e+03  
 Score: 39.00  
 Percent Similarity: 88.89%  
 Best Local Similarity: 88.89%  
 Query Match: 3  
 DB: 3

US-10-617-443B-2\_COPY\_1\_9 (1-9) x AC074202 (1-35546)

QY 1 MethiShleuAlaUpProAlaSerAla 9  
 DB 966 ATGACCTGGCCACCTGCATCGCT 992

RESULT 3  
 BX571965\_13  
 WPCOMENT

Sequence split into 41 fragments LOCUS BX571965 Accession BX571965  
 Fragment Name Begin End  
 BX571965\_00 1 110000  
 BX571965\_01 100001 210000  
 BX571965\_02 200001 310000  
 BX571965\_03 300001 410000  
 BX571965\_04 400001 510000  
 BX571965\_05 500001 610000  
 BX571965\_06 600001 710000  
 BX571965\_07 700001 810000  
 BX571965\_08 800001 910000  
 BX571965\_09 900001 1010000  
 BX571965\_10 1000001 1110000  
 BX571965\_11 1100001 1210000  
 BX571965\_12 1200001 1310000  
 BX571965\_13 1300001 1410000  
 BX571965\_14 1400001 1510000  
 BX571965\_15 1500001 1610000  
 BX571965\_16 1600001 1710000  
 BX571965\_17 1700001 1810000  
 BX571965\_18 1800001 1910000  
 BX571965\_19 1900001 2010000  
 BX571965\_20 2000001 2110000  
 BX571965\_21 2100001 2210000  
 BX571965\_22 2200001 2310000  
 BX571965\_23 2300001 2410000  
 BX571965\_24 2400001 2510000  
 BX571965\_25 2500001 2610000  
 BX571965\_26 2600001 2710000

len = 35546  
 29/27 = 1.074  
 27/29 = 0.931

Alignment Scores:  
Pred. No.: 1.04e+04 Length: 110000  
Score: 39.00 Matches: 8  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 1  
Query Match: 88.64% Indels: 0  
DB: 1 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x BX571965\_13 (1-110000)

QY 1 MethisLeuAlaLeuProAlaSerAla 9  
DB 53646 ATGCATCTGCACGCCGCTCCGCG 53672

RESULT 4  
CP000010\_19/c

WPCOMMENT Sequence split into 36 fragments LOCUS CP000010 Accession CP000010

Fragment Name	Begin	End
CP000010_00	1	110000
CP000010_01	100001	210000
CP000010_02	200001	310000
CP000010_03	300001	410000
CP000010_04	400001	510000
CP000010_05	500001	610000
CP000010_06	600001	710000
CP000010_07	700001	810000
CP000010_08	800001	910000
CP000010_09	900001	1010000
CP000010_10	1000001	1110000
CP000010_11	1100001	1210000
CP000010_12	1200001	1310000
CP000010_13	1300001	1410000
CP000010_14	1400001	1510000
CP000010_15	1500001	1610000
CP000010_16	1600001	1710000
CP000010_17	1700001	1810000
CP000010_18	1800001	1910000
CP000010_19	1900001	2010000
CP000010_20	2000001	2110000
CP000010_21	2100001	2210000
CP000010_22	2200001	2310000
CP000010_23	2300001	2410000
CP000010_24	2400001	2510000
CP000010_25	2500001	2610000
CP000010_26	2600001	2710000
CP000010_27	2700001	2810000
CP000010_28	2800001	2910000
CP000010_29	2900001	3010000
CP000010_30	3000001	3110000
CP000010_31	3100001	3210000
CP000010_32	3200001	3310000
CP000010_33	3300001	3410000
CP000010_34	3400001	3510000
CP000010_35	3500001	3510148

Continuation (20 of 36) of CP000010 from base 1900001 (CP000010 Burkholderia mallei ATCC

Alignment Scores:  
Pred. No.: 1.04e+04 Length: 110000  
Score: 39.00 Matches: 8  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 1  
Query Match: 88.64% Indels: 0  
DB: 1 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x CP000010\_19 (1-110000)

QY 1 MethisLeuAlaLeuProAlaSerAla 9  
DB 65822 ATGCATCTGCACGCCGCTCCGCG 65796

RESULT 5

AC087796/c AC087796 117146 bp DNA linear HTG 14-APR-2003  
LOCUS Leishmania major chromosome 35 clone P568 strain Friedlin, \*\*\*  
DEFINITION SEQUENCING IN PROGRESS \*\*\*

ACCESSION AC087796  
VERSION AC087796.4 GI:18158370  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Leishmania major  
ORGANISM Leishmania major  
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 117146)

AUTHORS Myler,P.J., Sisk,B., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,  
McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.

TITLE Direct Submission

JOURNAL Submitted (25-JAN-2001) Seattle Biomedical Research Institution, 4  
Nickerson Street, Seattle, WA 98109-1651, USA

REFERENCE 2 (bases 1 to 117146)

AUTHORS Myler,P.J., Sisk,B., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,  
McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.

JOURNAL Direct Submission  
Submitted (14-APR-2003) Seattle Biomedical Research Institution, 4  
Nickerson Street, Seattle, WA 98109-1651, USA

COMMENT On Jan 16, 2002 this sequence version replaced gi:15778699.  
\* This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 117146: contig of 117146 bp in length.  
FEATURES  
source 1. 117146  
location/Qualifiers  
/organism="Leishmania major"  
/mol\_type="genomic DNA"  
/strain="Friedlin"  
/db\_xref="taxon:5664"  
/chromosome="35"  
/clone="P568"

ORIGIN

Alignment Scores:

Pred. No.: 1.1e+04 Length: 117146  
Score: 39.00 Matches: 8  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 1  
Query Match: 88.64% Indels: 0  
DB: 2 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x AC087796 (1-117146)

QY 1 MethisLeuAlaLeuProAlaSerAla 9  
DB 29532 ATGCACCTGCGCACCGCTCCGCT 29506

RESULT 6  
AC118669/c  
LOCUS  
DEFINITION  
AC118669  
AC118669  
AC118669.2  
GI:22711575  
HTG.  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Eriarthroideae; Oryzoideae; Oryza.  
1 (bases 1 to 142401)  
McCombie, W.R., de la Bastide, M., Spiegel, L., Preston, R.,  
Nascimento, L., Zutavern, T., Balija, V., Bell, M., Miller, B.,  
Katzengraber, F., Muller, S., Sullivan, P., Yang, C., Dike, S.,  
O'Shaughnessy, A., Palmer, L. and Dedhia, N.  
Genomic sequence for Oryza sativa, Nipponbare strain, clone  
OSJNB0011G21, from chromosome 3, complete sequence  
Unpublished  
2 (bases 1 to 142401)  
McCombie, W.R.  
Direct Submission  
Submitted (20-APR-2002) Lita Annenberg Hazen Genome Center, Cold  
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,  
NY 11724, USA  
3 (bases 1 to 142401)  
McCombie, W.R.  
Direct Submission  
Submitted (04-SEP-2002) Lita Annenberg Hazen Genome Center, Cold  
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,  
NY 11724, USA  
On Sep 4, 2002 this sequence version replaced gi:20219081.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The nucleotide  
sequence of this BAC clone was generated by combining Syngenta and  
Cold Spring Harbor Laboratory Genome Center sequencing data.  
Clone OSJNB0011G21 overlaps clone OSJNB0085H08 (AC118340) from  
base 67378 to base 142401. The overlap is from base 52004 to base  
126708 (complemented) on OSJNB0085H08.  
Location/Qualifiers  
1. 142401  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="3"  
/clone="OSJNB0011G21"  
/clone\_1ib="Ecor1"  
11585\_11608  
/note="We believe the assembly to be correct. The  
sequence is a mononucleotide (C) repeat in which the exact  
number of Cs is unknown. One subclone spans the repeat  
into unique sequence on both sides."  
66157\_66177  
/note="We believe the assembly to be correct. The  
sequence is a mononucleotide (G) repeat in which the exact  
number of Gs is unknown. Four subclones in the region  
show one additional G compared to that represented by the  
assembly."

Query Match: 88.64% Indels: 0  
DB: 8 Gaps: 0  
US-10-617-443b-2\_COPY\_1\_9 (1-9) x AC118669 (1-142401)  
Ory 1 MethisLeuAlaLeuProAlaSerAla 9  
Db 33763 CTCACCTACGACATGCCGCTCCGGC 33737  
RESULT 7  
AC144913  
LOCUS  
DEFINITION  
AC144913  
AC144913.2  
GI:33086398  
HTG; HTGS PHASST; HTGS DRAFT.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 150106)  
Birren, B., Nussbaum, C. and Lander, E.  
Unpublished  
2 (bases 1 to 150106)  
Birren, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, S., Arachchi, H.M., Barna, N., Baetien, V., Bloom, T.,  
Bogunlavsky, L., Boukhalter, B., Canarata, J., Chang, J., Choepel, Y.,  
Colymore, A., Cook, A., Cooke, P., Cornu, B., Dearellano, K.,  
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
Ferreira, P., Fitzgerald, M., Gage, D., Galsgan, J., Gardyna, S.,  
Graham, L., Grand-Pierre, N., Hatz, N., Haggopian, D., Hagos, B.,  
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,  
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Rachpka, A., Ramasamy, U., Raymond, C., Retta, R., Riese, C., Rogov, P.,  
Roman, J., Schauer, S., Schupack, R., Seaman, S., Seery, P., Smith, C.,  
Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (25-MAY-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 150106)  
Birren, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, M., Arachchi, H.M., Barna, N., Baetien, V., Bloom, T.,  
Bogunlavsky, L., Boukhalter, B., Canarata, J., Chang, J., Choepel, Y.,  
Colymore, A., Cook, A., Cooke, P., Cornu, B., Dearellano, K.,  
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
Ferreira, P., Fitzgerald, M., Gage, D., Galsgan, J., Gardyna, S.,  
Graham, L., Grand-Pierre, N., Hatz, N., Haggopian, D., Hagos, B.,  
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C.,  
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Rachpka, A., Ramasamy, U., Raymond, C., Retta, R., Riese, C., Rogov, P.,  
Roman, J., Schauer, S., Schupack, R., Seaman, S., Seery, P., Smith, C.,  
Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (21-JUL-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA



Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,  
 Ferreira, S., Friese, E., Galle, R.F., Garg, N.S., George, R.A.,  
 Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,  
 Ileguam, C., Jalali, M., Kruse, D., Li, P., Matea, B., Moshrefi, A.,  
 McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,  
 Pacle, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,  
 Phouanavong, S., Piltman, G.S., Puri, V., Richarde, S., Scheeler, F.,  
 Stapleton, M., Strong, R., Svirekas, R., Tector, C., Williams, S.M.,  
 Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.  
 Sequencing of Drosophila chromosome 2R, region 56A2-56B1  
 Unpublished  
 2 (bases 1 to 160650)  
 Celniker, S.E., Agbayani, A., Arcana, T.T., Baxter, E., Blazey, R.G.,  
 Buttenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
 Hoskins, R.A., Houston, K.A., Hummachi, S.R., Katta, K., Kearney, L.,  
 Kim, E., Lee, B., Lewis, S., Li, P., Lomutan, M.A., Mazda, P.,  
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S.,  
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,  
 Svirekas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
 Rubin, G.M.  
 Direct Submission  
 Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley  
 Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 On Apr 27, 2001 this sequence version replaced gi:6598785.  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720  
 This sequence was assembled using end sequences from a whole genome  
 shotgun and from subclones of this BAC and its neighboring clones.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdgs@fruitfly.berkeley.edu](mailto:bdgs@fruitfly.berkeley.edu).  
 Location/Qualifiers  
 1. 160650  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="2R"  
 /map="56A2-56B1"  
 /clone="BACR08118 (D660)"  
 /clone\_lib="RP1-98 (Rowell) Park Cancer Institute  
 Drosophila melanogaster BAC library, partial EcoRI in  
 BAC3.6)"

## ORIGIN

## Alignment Scores:

Pred. No.: 1.46e+04 Length: 160650  
 Score: 39.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 88.64% Indels: 0  
 DB: 3 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x ACO08188 (1-160650)

QY 2 H1sleu1aleuPro1a1aSer1a 9  
 Db 81657 CACTTGCGATTGCGACGCCGCC 81634

RESULT 9  
 ACO18184 160885 bp DNA linear HTG 09-DEC-1999  
 LOCUS Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 DEFINITION ACO18184  
 ACCESSION ACO18184 GI:6553007  
 VERSION ACO18184.1  
 KEYWORDS HTG; HTGS\_PHASE2.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 160885)  
 AUTHORS Adams, M. and Venter, J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 COMMENT This sequence was identified as CDW:10213716 by the submitter.  
 For more information on this record e-mail to [fly@celera.com](mailto:fly@celera.com).  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 Location/Qualifiers  
 1. 160885  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 1. 167588  
 /organism="Oryza sativa"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4530"  
 /chromosome="3"  
 /clone="OSUNBa0016B07"

## ORIGIN

## Alignment Scores:

Pred. No.: 1.47e+04 Length: 160885  
 Score: 39.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 88.64% Indels: 0  
 DB: 2 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x ACO18184 (1-160885)

QY 2 H1sleu1aleuPro1a1aSer1a 9  
 Db 125771 CACTTGCGATTGCGACGCCGCC 125748

RESULT 10  
 AF461424 167588 bp DNA linear PLN 26-JUN-2003  
 LOCUS Oryza sativa chromosome 3 clone OSUNBa0016B07, complete sequence.  
 DEFINITION AF461424  
 ACCESSION AF461424.2 GI:32265039  
 VERSION AF461424.2  
 KEYWORDS HTG.  
 SOURCE Oryza sativa  
 ORGANISM Oryza sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 167588)  
 AUTHORS Eastman, A.P., Smith, S.C., Gingle, A.R., Pratt, L.H. and  
 Cordomier-Pratt, M.-M.  
 Cordonier-Pratt, M.-M.

## REFERENCE

## AUTHORS

## TITLE

JOURNAL Submitted (18-DEC-2001) Botany, University of Georgia, 3604 Miller  
 Plant Sciences, Athens, GA 30606, USA  
 2 (bases 1 to 167588)  
 REFERENCES Eastman, A.P., Smith, S.C., Berlin, N., Liang, C., Gingle, A.R.,  
 Pratt, L.H. and Cordomier-Pratt, M.-M.

## AUTHORS

## TITLE

## JOURNAL

Submitted (26-JUN-2003) Botany, University of Georgia, 3604 Miller  
 Plant Sciences, Athens, Georgia 30606, USA  
 Sequence update by submitter  
 On Jun 26, 2003 this sequence version replaced gi:18150816.

## REMARK

## COMMENT

## FEATURES

## source

## ORIGIN

## Alignment Scores:

Pred. No.: 1.52e+04 Length: 167588  
 Score: 39.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2

Beet Local Similarity: 77.78% Mismatches: 0  
 Query Match: 88.64% Indels: 0  
 DB: 8 Gaps: 0

US-10-617-443b-2\_COPY\_1\_9 (1-9) x APr61424 (1-167588)

QY 1 Mech1aLeu1aLeuPro1aSer1a 9  
 ::::::::::::::::::::::::::::::

Db 12713 CTCACACTAGCAGTCCCGCTCCGCG 12739

RESULT 11  
 AC107764 196178 bp DNA linear ROD 07-JUL-2004  
 LOCUS Mus musculus chromosome 15, clone RP23-179A6, complete sequence.  
 DEFINITION AC107764  
 AC107764.11 GI:49745564  
 VERSION  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Birren, B., Nussbaum, C. and Lander, E.  
 JOURNAL 1 (bases 1 to 196178)  
 REFERENCE 2 (bases 1 to 196178)  
 AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Baetien, V., Boguslavsky, L., Bouhgalter, B.,  
 Brown, A., Camarata, J., Campolino, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Coyette, M., Graham, L., Grand-Pierre, N.,  
 Hager, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,  
 Landers, T., Lehoczeky, J., Levine, R., Liu, G., Maclean, C.,  
 Macdonald, P., Major, J., Margis, N., Matthews, C., McCarthy, M.,  
 McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,  
 Mlenaga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
 Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,  
 Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,  
 Strause, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 196178)

TITLE Direct Submission  
 JOURNAL Submitted (11-MAY-2004) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 4 (bases 1 to 196178)  
 AUTHORS Birren, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, S., Barna, N., Baetien, V., Boguslavsky, L., Bouhgalter, B.,  
 Bloom, T., Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J.,  
 Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,  
 Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L.,  
 Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D.,  
 Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hager, N.,  
 Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,  
 Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
 Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,  
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenaga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunhkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,  
 Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R.,  
 Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N.,  
 Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R.,  
 Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,  
 Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (07-JUL-2004) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 7, 2004 this sequence version replaced gi:47106120.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE  
 JOURNAL  
 COMMENT  
 Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: WIRB  
 Web site: http://www.seq.wi.mit.edu  
 Contact: sequence submissions@broad.mit.edu  
 Project Information  
 Center project name: 120413  
 Center clone name: 179\_A6

FEATURES  
 source  
 Location/Qualifiers  
 1. 196178  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="15"  
 /map="15"  
 /clone="RP23-179A6"  
 /clone\_1lb="RP21-23 Female Mouse BAC"  
 1. 84  
 /rpt\_family="B1\_MM"  
 1. 6  
 /note="clone boundary  
 clone end:SP6  
 site:BCOR1"  
 102. 213  
 /rpt\_family="PB1D7"  
 214. 255  
 /rpt\_family="PB1D7"  
 265. 428  
 /rpt\_family="B3"  
 429. 475  
 /rpt\_family="PB1D7"  
 1982. 2027  
 /rpt\_family="CNAAN)n"  
 complement(3183..3312)  
 /rpt\_family="B1F"  
 7340. 7423  
 /note="single clone coverage"  
 8810. 8925  
 /rpt\_family="B1F"  
 9195. 9262  
 unsure

```

/nc="single clone coverage"
repeat_region 10007.10115 /rpt family="B1_MM"
complement(112227.11336)
repeat_region /rpt family="PBID9"
complement(11361.11471)
repeat_region /rpt family="MIR"
11798.11847
/rpt family="MIR3"
repeat_region complement(12475.12582)
/rpt family="B1_MM"
repeat_region complement(12632.12760)
/rpt family="B1_MM"
repeat_region complement(13035.13137)
/rpt family="PBID7"
13396.13475
/rpt family="CATATA)n"
13671.13744
/rpt family="TA)n"
13745.13764
/rpt family="TG)n"
repeat_region complement(13815.13903)
/rpt family="PBID7"
repeat_region complement(13830.13922)
/rpt family="ID_B1"
repeat_region complement(14135.14866)
/rpt family="LIME2"
complement(14902.15095)
repeat_region /rpt family="B2_MM2"
15162.15213
/rpt family="TATG)n"
complement(15312.15477)
repeat_region /rpt family="RSINE1"
complement(15478.15748)
/rpt family="LIME2"
repeat_region complement(15779.15864)
/rpt family="B1_MM"
repeat_region complement(15895.16043)
/rpt family="L3"
16082.16220
/rpt family="B1_MM"
repeat_region complement(16327.16466)
/rpt family="B1_MM"
repeat_region complement(16615.16735)
/rpt family="B1_MM"
repeat_region complement(16824.16965)
/rpt family="B3"
16968.16991
/rpt family="AT_rich"
repeat_region complement(17450.17553)
/rpt family="PBID10"
repeat_region complement(18238.18358)
/rpt family="ID_B1"
21795.21898
/rpt family="PBID10"
repeat_region complement(23074.23267)
/rpt family="B3A"
repeat_region complement(23435.23562)
/rpt family="B1_MM"
repeat_region complement(23437.23598)
/rpt family="B4"
complement(23708.23884)
/rpt family="B4A"
24013.24050
/rpt family="CTG)n"
repeat_region 24287.24322

```

```

Alignment Scores:
Pred. No.: 1.76e+04 Length: 196178
Score: 39.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 88.64% Indels: 0

```

```

DB: 10 Gaps: 0
US-10-617-443b-2_COPY_1_9 (1-9) x AC107764 (1-196178)
QY 2 HisleuAlaLeuProAlaSerAla 9
Db 263 CATCTTCTCTCTCTGCTCGGCC 240

RESULT 12
AC146955/c
LOCUS
DEFINITION
Ocolemur garnettii clone CH256-263M9, WORKING DRAFT SEQUENCE, 16
unordered pieces.
ACCESSION
AC146955.1 GI:38044153
VERSION
AC146955
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Ocolemur garnettii (small-eared galago)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Strepsirrhini; Galagonidae; Ocolemur.
1 (bases 1 to 218827)
Antoniellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
Coleman,H., Engle,J., Granice,S., Guan,X., Gupta,J., Haghighi,P.,
Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R.,
Karins,E., Kwong,P., Latic,P., Larson,S., Lee-Jin,S.-O.,
Legaspi,R., Maduro,Q.U., Maduro,V.B., Marquis,B.H., Mastello,C.,
Maskeki,B., McDowell,J., Mullikin,J.C., Pegurigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schandler,K.,
Schueler,M.G., Shah,K., Sison,C., Stantropop,S., Thomas,J.W.,
Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and
Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 218827)
Green,E.D.
Direct Submission
Submitted (30-OCT-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gathersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: fkv
Center clone name: 263M09
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 213879 bases at least Q40
Consensus quality: 215076 bases at least Q30
Consensus quality: 215592 bases at least Q20
Insert size: 23400; agarose-fp
Insert size: 217327; sum-of-contigs
Quality coverage: 9.25x in Q20 bases; sum-of-contigs
Quality coverage: 9.96x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2402: contig of 2402 bp in length
* 2403 2502: gap of unknown length
* 2503 2514: contig of 7012 bp in length
* 2515 9614: gap of unknown length
* 9615 14216: contig of 4602 bp in length
* 14217 14316: gap of unknown length

```



```

* 14317 24704: contig of 10388 bp in length
* 24705 24804: gap of unknown length
* 24805 31640: contig of 6836 bp in length
* 31641 31740: gap of unknown length
* 31741 43787: contig of 12047 bp in length
* 43788 43887: gap of unknown length
* 43888 51985: contig of 8097 bp in length
* 51985 52084: gap of unknown length
* 52085 63590: contig of 11506 bp in length
* 63591 63690: gap of unknown length
* 63691 73810: contig of 10120 bp in length
* 73811 73910: gap of unknown length
* 73911 88004: contig of 14094 bp in length
* 88005 88104: gap of unknown length
* 88105 107084: contig of 18980 bp in length
* 107085 107184: gap of unknown length
* 107185 124420: contig of 17136 bp in length
* 124421 124421: gap of unknown length
* 124421 145411: contig of 20991 bp in length
* 145412 145511: gap of unknown length
* 145512 164902: contig of 19391 bp in length
* 164903 165002: gap of unknown length
* 165003 189267: contig of 24265 bp in length
* 189268 189367: gap of unknown length
* 189368 218827: contig of 29460 bp in length.

```

```

FEATURES
source
1. .218827
/organism="Orolemur garnettii"
/mol_type="genomic DNA"
/db_xref="taxon:30611"
/clone="CH256-263M9"
/clone_1b="CH256"
1. .2402

```

```

misc_feature /note="assembly_fragment" 2503..9514
misc_feature /note="assembly_fragment" 9615..14216
misc_feature /note="assembly_fragment" 14317..24704
misc_feature /note="assembly_fragment" 24805..31640
misc_feature /note="assembly_fragment" 31741..43787
misc_feature /note="assembly_fragment" 43888..51984
misc_feature /note="assembly_fragment" 52085..63590
misc_feature /note="assembly_fragment" 63691..73810
misc_feature /note="assembly_fragment" 73911..88004
misc_feature /note="assembly_fragment" 88105..107084
misc_feature /note="assembly_fragment" 107185..124420
misc_feature /note="assembly_fragment" 124421..145411
misc_feature /note="assembly_fragment" 145512..164902
misc_feature /note="assembly_fragment" 165003..189267
misc_feature /note="assembly_fragment" 189368..218827
misc_feature /note="assembly_fragment"

```

## ORIGIN

```

Alignment Scores:
Prod. No.: 1.94e+04 Length: 218827
Score: 39.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 88.64% Indels: 0
DB: 2 Gaps: 0

```

```

US-10-617-443B-2_COPY_1_9 (1-9) x AC146955 (1-218827)
Oy
2 HsleuAla1euproAlaSerAla 9
Db 182222 CATCTGCTCTTCACGCTCTGTC 182199

```

```

RESULT 13
AC137178/c
LOCUS
DEFINITION
Rattus norvegicus clone CH250-unknown, *** SEQUENCING IN PROGRESS
***, 10 unordered pieces.
AC137178 221049 bp DNA linear HTG 19-NOV-2002
Rattus norvegicus

```

```

AC137178 1 GI:25073057
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM

```

## REFERENCE

## AUTHORS

1 (bases 1 to 221049)

Wuzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, W., Benahmed, F., Biewalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Dudkin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Georige, G., Georjia, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hernandez, J., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., L., Hodgson, A., Hughes, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, R., Johnson, R., Jollivet, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kovar, C., Kapatly, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenschewa, L., Louised, H., Lozada, R., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naik, L., Nankervill, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwackelam, O., Okmonu, G., Olarnpunaagoon, A., Pal, S., Parke, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Richards, S., Ritz, S., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, M., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Smjs, D., Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Soia, J., Stearle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tinsley, A., Trejos, Z., Uemari, K., Valas, R., Vera, V., Villaseana, D., Waldon, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, K., Williams, G., Willson, R., Wlezyk, R., Woden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, D., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## TITLE

## JOURNAL

## REFERENCE

2 (bases 1 to 221049)





\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 102058: contig of 102058 bp in length  
 \* 102059 102158: gap of 100 bp  
 \* 102159 113284: contig of 11126 bp in length  
 \* 113285 113384: gap of 100 bp  
 \* 113385 121021: contig of 7637 bp in length  
 \* 121022 121121: gap of 100 bp  
 \* 121122 156867: contig of 35746 bp in length  
 \* 156868 162929: contig of 5962 bp in length  
 \* 162930 163030: gap of 100 bp  
 \* 163030 184202: contig of 21073 bp in length  
 \* 184202 184203: gap of 100 bp  
 \* 184203 221386: contig of 37384 bp in length.

## FEATURES

source  
 1. 221586  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone\_DKEY="170K13"  
 /clone\_idb="Daniokey"  
 1. 102058  
 /note="assembly\_fragment:02700  
 fragment\_chain:1"  
 102159 113284  
 /note="assembly\_fragment:00743  
 fragment\_chain:1"  
 113385 121021  
 /note="assembly\_fragment:00121  
 fragment\_chain:1"  
 121122 156867  
 /note="assembly\_fragment:00225  
 fragment\_chain:1"  
 156868 162929  
 /note="assembly\_fragment:00123  
 fragment\_chain:1"  
 163030 184102  
 /note="assembly\_fragment:01032  
 fragment\_chain:1"  
 184203 221586  
 /note="assembly\_fragment:01392  
 fragment\_chain:1"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,966+04 Length: 221586  
 Score: 39.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 88.64% Indels: 0  
 DB: 2 Gaps: 0

US-10-617-443b-2\_COPY\_1\_9 (1-9) x BX957332 (1-221586)

Qy 2 H18LEUAlaleuPROAlaserAla 9

Db 167230 CATCTTGCCCTGCGGCTTCAGCC 167253

## RESULT 15

AC099140 238103 bp DNA linear HTG 10-MAY-2003

LOCUS Rattus norvegicus clone CH230-52A14, WORKING DRAFT SEQUENCE, 4

DEFINITION unorderd pieces.

AC099140 AC099140.12 GI:30522132

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

## SOURCE

## ORGANISM

Rattus norvegicus (Norway rat)

## REFERENCE

## AUTHORS

Wuzny,D,Marie, Metzker,M,lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alebrooks,S, Amn,A, Angiano,D, Anyalebech,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benamed,F, Bialwal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhey,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Crete,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Dlyva,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Evans,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Frazer,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunaratne,P, Healand,W, Hamil,C, Hamilton,C, Hamilton,K, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huliyk,S, Hume,J, Iglebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levay,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorensuhewa,L, Louisseged,H, Lozada,R,J, Lu,X, Me,J, Maheshwari,M, Mahindartne,M, Mahmud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Matinez,E, McInerney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Miosavilevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwakoelamoh,O, Okwuonu,G, Olariunsgoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Plannoch,C, Plopper,F, Poindexter,A, Popovic,D, Prilina,E, Pu,L, L, Rives,C, Rodkey,T, Rojase,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,M, Savery,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sittler,C,D, Sosa,J, Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Taylor,C, Taylor,T, Thomas,N, Thomas,A, Syatek,A, Taber,P, Tabor,C, Teagle,M, Strong,R, Sutton,A, Syatek,A, Tingey,Z, Usmani,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Wilson,R, Wiczek,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,S, Zhao,S, Dunn,D, von Weidenhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R.A.

## TITLE

## JOURNAL

## AUTHORS

## REFERENCE

## TITLE

## JOURNAL

## COMMENT

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

#### Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: GGYK

Center clone name: CH230-52A14

Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 232295 bases at least Q40

Consensus quality: 232971 bases at least Q30

Estimated insert size: 241423; sum-of-contigs estimation

Quality coverage: 14x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      234672: contig of 234672 bp in length
*      234673 234772: gap of unknown length
*      234773 235808: contig of 1036 bp in length
*      235809 235909: gap of unknown length
*      235910 236913: contig of 1005 bp in length
*      236914 237013: gap of unknown length
*      237014 238103: contig of 1090 bp in length.
Location/Qualifiers
1. .238103
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-52A14"
1. .1535
/notes="wgs end extension
clone_end:17"
1634. .2403
/notes="clone_boundary
clone_end:17
site:BCORI
end_sequence:BH344407"
complement(232158..232850)
/notes="clone_boundary
clone_end:5p6
site:BCORI
end_sequence:BH344409"

```

#### FEATURES

##### source

##### misc\_feature

##### misc\_feature

##### misc\_feature

#### ORIGIN

##### Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

2.09e+04 Length: 238103

39.00 Matches: 8

100.00% Conservative: 0

100.00% Mismatches: 0

88.64% Indels: 0

2 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x AC099140 (1-238103)

Qy 2 HistenuAlaleuproAlasera1a 9  
 Db 172836 CATTCGCCCTCCCTGCTTCGCC 172859

Search completed: June 29, 2005, 23:51:05  
 Job time : 337.461 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 29, 2005, 18:37:40 ; Search time 25.2682 Seconds

(without alignment)  
2108.485 Million cell updates/sec

Title: US-10-617-443B-2\_COPY\_1\_9  
Perfect score: 44  
Sequence: 1 MHLALPASA 9

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 segs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seg length: 0  
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=x1h  
-Q=/cgn2\_1/USPto.spool/US10617443/runat.29062005.135931.26052/app.query.fasta\_1.718  
-DB=N\_Geneseq.16Dec04 -QPM=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdd  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10617443 @CGN\_1\_1-556 @runat.29062005.135931.26052 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGOUDRY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq.16Dec04:\*  
1: geneeqn1980s:\*  
2: geneeqn1990s:\*  
3: geneeqn2000s:\*  
4: geneeqn2001as:\*  
5: geneeqn2001bs:\*  
6: geneeqn2002as:\*  
7: geneeqn2002bs:\*  
8: geneeqn2003as:\*  
9: geneeqn2003bs:\*  
10: geneeqn2003cs:\*  
11: geneeqn2003ds:\*  
12: geneeqn2004as:\*  
13: geneeqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	3006	12	ADJ11355 Human PRS
2	39	88.6	754	6	ABST77428
3	39	88.6	963	8	ACA25538
4	39	88.6	2157	8	ACA25970
5	38	86.4	578	10	ADG37683

C	6	38	86.4	60940	9	ADA02582	Ada02582 Human REL
C	7	38	86.4	60940	10	ADB72320	ADB72320 Human REL
C	8	38	86.4	60940	10	ADB58830	ADB58830 Human REL
C	9	38	86.4	110000	4	AA199682_19	Continuation (20 o
C	10	37	84.1	2196	2	AAQ67901	Aaq67901 Syndecan
C	11	37	84.1	2196	2	AAV41533	Aav41533 Nucleotid
C	12	37	84.1	2196	2	AAV15947	Aav15947 Mouse syn
C	13	37	84.1	2196	8	ACA60751	ACA60751 Mouse syn
C	14	37	84.1	21562	11	ACNA45004	ACNA45004 Mouse gen
C	15	37	84.1	30393	4	AAK67239	AAK67239 Human imm
C	16	37	84.1	247544	12	ADQ59419	Adq59419 Human can
C	17	36	81.8	34	12	ADJ11359	Adj11359 PCR prime
C	18	36	81.8	706	9	ACT4588	Act4588 DNA clone
C	19	36	81.8	756	8	ACA22808	Aca22808 Prokaryot
C	20	36	81.8	757	8	ACA21959	Aca21959 Prokaryot
C	21	36	81.8	795	12	ADN62510	Adn62510 A. thalia
C	22	36	81.8	1077	3	AAC40009	Aac40009 Arabidops
C	23	36	81.8	1387	8	ACA57488	Aca57488 Human adi
C	24	36	81.8	1395	13	ADQ85511	Adq85511 Human tum
C	25	36	81.8	1797	12	ADQ64218	Adq64218 Novel hum
C	26	36	81.8	1854	4	AA159099	Aa159099 Human pol
C	27	36	81.8	1854	4	AA160885	Aa160885 Human pol
C	28	36	81.8	1854	5	ADQ93322	Adq93322 DNA encod
C	29	36	81.8	1854	9	ADB49082	Adb49082 Novel hum
C	30	36	81.8	1939	6	ABA01852	Abao1852 Human reg
C	31	36	81.8	2057	10	AD121987	Ad121987 Novel hum
C	32	36	81.8	2160	9	ADA27221	Ada27221 Human NOV
C	33	36	81.8	2203	10	ADB95581	Adb95581 Human NOV
C	34	36	81.8	2250	12	ADO22235	Ado22235 Human WNT
C	35	36	81.8	2359	10	ADB58551	Adb58551 Human gen
C	36	36	81.8	2374	4	AAK78587	Aak78587 Human imm
C	37	36	81.8	2389	8	ACC48788	Acc48788 Human Evi
C	38	36	81.8	2389	8	ACC48792	Acc48792 Human Evi
C	39	36	81.8	2491	10	ADA53235	Ada53235 Human cod
C	40	36	81.8	2566	10	ADD49028	Add49028 Human NOV
C	41	36	81.8	2569	3	AAC93383	Aac93383 Human sec
C	42	36	81.8	2590	6	ABN85556	Abn85556 Human lat
C	43	36	81.8	2629	5	ADB95583	Adb95583 Human NOV
C	44	36	81.8	2707	5	AA544925	Aa544925 CDNA enco
C	45	36	81.8	2714	10	ADA53450	Ada53450 Human cod

# ALIGNMENTS

RESULT 1	ADJ11355	ADJ11355 standard; CDNA; 3006 BP.
XX	ADJ11355;	
AC	ADJ11355;	
XX	20-MAY-2004	(first entry)
DT	20-MAY-2004	(first entry)
XX	Human PRSS11-L	CDNA encoding a serine protease Segid 1.
DE	Human PRSS11-L	CDNA encoding a serine protease Segid 1.
XX	human, PRSS11-L; gene; ser; s2 serine protease; s2/HTRA;	
KW	cellular physiology; apoptosis; gene therapy; osteoarthritis; stress;	
KW	apoptotic; osteopathic; antiarthritic; tranquiliser.	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	Key	Location/Qualifiers
FT	CDS	1011..2015
FT		/tag= a
FT		/product= "PRSS11-L protein"
XX	US2004005659-A1.	
XX	08-JAN-2004.	
XX	03-JUL-2002;	2002US-00189099.
XX	03-JUL-2002;	2002US-00189099.
XX	03-JUL-2002;	2002US-00189099.

PA (DARR/) DARROW A L.  
 PA (OLJ/) OL J.  
 PA (CHEN/) CHEN C.  
 PA (ANDR/) ANDRADE-GORDON P.  
 XX Darrow AL, OL J, Chen C, Andrade-Gordon P;  
 XX WPI; 2004-081723/08.  
 DR P-PSDB; ADJ11356.  
 XX  
 XX  
 PT New isolated S2 serine protease nucleic acids and polypeptides, useful  
 PT for diagnosing and/or treating diseases with aberrant expression or  
 PT activity the S2 serine protease, such as osteoarthritis, stress and  
 PT apoptotic disorders.  
 XX  
 PS Claim 2; SEQ ID NO 1; 28pp; English.  
 XX  
 CC This invention relates to a novel isolated nucleic acid molecule  
 CC identified as PRSS11-L that encodes an S2 serine protease. Specifically,  
 CC it refers to members of the S2/Htra serine protease family, such that it  
 CC plays a role in cellular physiology and apoptosis. The present invention  
 CC provides agonists, antagonists, antibodies and recombinant expression  
 CC vectors useful in methods of treatment, or detection and diagnosis of  
 CC diseases associated with the aberrant expression or activity of the S2  
 CC serine protease, PRSS11-L. Accordingly, compositions described herein can  
 CC be used via gene therapy routes to treat osteoarthritis, stress and  
 CC apoptotic disorders. As such, they exhibit osteopathic, antiarthritic and  
 CC tranquiliser activities. This polynucleotide sequence is the human PRSS11  
 CC -L cDNA sequence of the invention.  
 XX  
 SQ Sequence 3006 BP; 615 A; 918 C; 911 G; 562 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 166 Length: 3006  
 Score: 44.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 12 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x ADJ11355 (1-3006)

Qy 1 Mech1sleuAlaleuProAlaSerAla 9  
 Db 1011 ATGCACCTGGCCTTCCCGCCAGCGCA 1037

RESULT 2  
 ABS77428  
 ID ABS77428 standard; cDNA; 754 BP.  
 XX  
 XX ABS77428;

DT 12-DEC-2002 (first entry)

DE Frog embryonic gene sequence Q9925836.

KW Frog; ss; embryonic development; developmental disorder; microarray;  
 KW cell differentiation.

OS Xenopus laevis.

PN US2002081610-A1.

PD 27-JUN-2002.

PF 23-JUL-2001; 2001US-00910943.

PR 21-JUL-2000; 2000US-0219558P.

PA (UYRQ ) UNIV ROCKEFELLER.

PI Hemmati-Briyvanlou A, Altman CR;

DR WPI; 2002-626534/67.

XX Nucleic acid array containing Xenopus embryonic nucleic acids is useful  
 PT to identify genes involved in embryonic development, to identify  
 PT different types of embryonic cells, and to diagnose developmental  
 PT disorders.

PS Claim 1; Page 698; 823pp; English.

XX  
 XX  
 CC The invention relates to a nucleic acid array, where each coordinate  
 CC contains a single nucleic acid species having one of 770 nucleotide  
 CC sequences (appearing as ABS76747-ABS77516) a of a Xenopus embryonic gene  
 CC product, or its complement or hybridisable fragment of not less than 20  
 CC contiguous nucleotides of one of those sequences. Also included are  
 CC detecting differential expression of embryonic genes, comprising: (a)  
 CC contacting a nucleic acid array comprising genes expressed in embryonic  
 CC but not mature cells with nucleic acids from sample and control cells;  
 CC and (b) detecting differential hybridisation of nucleic acids from the  
 CC sample cells relative to the control cells; and detecting defects in  
 CC development, comprising: (a) contacting nucleic acids from test cells  
 CC undergoing development with a nucleic acid array of gene products known  
 CC to play a fundamental role in the development process; and (b) detecting  
 CC a difference in expression of a fundamental gene in the sample cells  
 CC relative to a standard. The invention is useful to identify genes  
 CC involved in embryonic development and related processes such as cell  
 CC differentiation. This would be useful for diagnosing developmental  
 CC disorders and for identifying different types of embryonic cells. The  
 CC present sequence is one of the 770 Xenopus embryonic cDNA sequences  
 XX

SQ Sequence 754 BP; 169 A; 200 C; 171 G; 207 T; 0 U; 7 Other;

Alignment Scores:  
 Pred. No.: 346 Length: 754  
 Score: 39.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 88.64% Indels: 0  
 DB: 6 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x ABS77428 (1-754)

Qy 1 Mech1sleuAlaleuProAlaSerAla 9  
 Db 327 TTACATATCGCTTCGCGCCTCTGCT 353

RESULT 3  
 ACA25538  
 ID ACA25538 standard; DNA; 963 BP.  
 XX  
 XX ACA25538;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #7195.

KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.

OS Burkholderia fungorum.

PN WO20027183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

PA (ELIT-) ELITRA PHARM INC.

PI Wang L., Zamudio C., Malone C., Haselbeck R., Ohlsen K.L., Zykkind JW;  
PI Wall D., Trawick JD., Carr GJ., Yamamoto R., Forsyth RA., Xu HH;  
XX WPI; 2003-029926/02.  
DR P-PsDB; ABU21668.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 13408; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WHO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 963 BP; 176 A; 300 C; 322 G; 165 T; 0 U; 0 Other;

XX Alignment Scores:

	Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	452	963	8	0	0	0	0
Percent Similarity:	39.00						
Best Local Similarity:	88.89%						
Query Match:	88.64%						
Db:	8						

US-10-617-443B-2\_COPY\_1\_9 (1-9) x ACA25538 (1-963)

OY 1 Methistenevalaleuprolineaseria 9

Db 739 ATGCAATCGGCCACCGCGCTCGGGCG 765

RESULT 4  
ACA25970  
ACA25970 standard; DNA; 2157 BP.

AC ACA25970;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #7627.

XX Antisense, ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.

10x=CTG  
25/27 = 92.96%

XX Burkholderia mallei.  
OS  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002MO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
XX  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0349223P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyeck JM;  
PI Wall D, Traxwick JD, Carr GJ, Yamamoto R, Foreych RA, Xu HH;  
PI P-PSDB; ABU22100.  
DR WPI; 2003-029926/02.  
PT  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

Claim 14; SEQ ID NO 13840; 1766bp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are:  
(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the targets of prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 2157 BP; 350 A; 702 C; 761 G; 344 T; 0 U; 0 Other;

Alignment Scores:

	Score:	Length:
Percent Similarity:	1.08e+03	2157
Best Local Similarity:	39.00	Matches: 8
Query Match:	88.89%	Conservative: 0
	88.89%	Mismatches: 1
	88.64%	Indels: 0
		Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x ACA25970 (1-2157)





OS Homo sapiens.  
XX  
XX WO2003008583-A2.  
XX  
XX 30-JAN-2003.  
XX  
XX 26-DEC-2001; 2001WO-US051291.  
XX  
XX 02-MAR-2001; 2001US-00798586.  
XX 23-OCT-2001; 2001US-00004113.  
XX 08-NOV-2001; 2001US-00052482.  
XX 30-NOV-2001; 2001US-00997722.  
XX 20-DEC-2001; 2001US-00034650.  
XX  
XX (SAGR-) SAGRES DISCOVERY.  
XX  
XX Morris DW, Engelhard EK;  
XX  
XX WPI; 2003-239337/23.  
XX  
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
XX cancers, neoplasms, adenocarcinoma, or sarcomas.  
XX  
XX Claim 1; SEQ ID NO 148; 2304bp; English.  
XX  
XX The invention relates to a novel recombinant nucleic acid comprising a  
XX nucleotide sequence selected from any of the 660 sequences fully defined  
XX in the specification. A polynucleotide of the invention has cytostatic  
XX activity, and may have a use in gene therapy, or in a vaccine. The  
XX recombinant nucleic acids and polypeptides are useful for treating  
XX carcinomas, e.g. lymphomas, cancers, neoplasms, adenocarcinoma, and  
XX sarcomas. The present sequence represents a human gene of the invention.  
XX  
SQ Sequence 60940 BP; 16572 A; 11094 C; 10947 G; 18968 T; 0 U; 3359 Other;  
Alignment Scores:  
Pred. No.: 6.34e+04 Length: 60940  
Score: 38.00 Matches: 8  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 1  
Query Match: 86.36% Indels: 0  
DB: 10 Gaps: 0  
US-10-617-443b-2\_COPY\_1\_9 (1-9) x ADB72320 (1-60940)  
QY 1 MechisteleuAlaLeuProAlaSerAla 9  
DB 11468 ATGCATTAGCCCTCCCTGCTTGTCT 11442  
RESULT 8  
ADE95830/C  
ID ADE95830 standard; DNA; 60940 BP.  
XX  
XX ADE95830;  
XX  
XX 12-FEB-2004 (first entry)  
XX  
XX Human REL gene genomic DNA sequence.  
XX  
XX cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;  
XX lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; REL.  
XX  
XX Homo sapiens.  
XX  
XX WO2003039484-A2.  
XX  
XX 15-MAY-2003.  
XX  
XX 08-NOV-2002; 2002WO-US036071.  
XX  
XX 08-NOV-2001; 2001US-00052482.  
XX  
XX (SAGR-) SAGRES DISCOVERY.  
XX

XX  
XX Morris DW, Engelhard EK;  
XX  
XX WPI; 2003-441462/41.  
XX  
XX New carcinoma associated nucleic acids and proteins, useful for screening  
XX drug candidates, or for diagnosing and treating carcinomas, e.g.  
XX lymphoma, breast cancer, prostate cancer or leukemia.  
XX  
XX Claim 1; SEQ ID NO 88; 793bp; English.  
XX  
XX This invention relates to novel recombinant nucleic acids for use in  
XX diagnosis and treatment of cancer, especially carcinomas, as well as the  
XX use of compositions in screening methods. The compositions of the  
XX invention may have cytostatic activity whilst the disclosed sequences may  
XX be useful for gene therapy. The carcinoma associated nucleic acids and  
XX proteins are useful for diagnosing and treating carcinomas, for example  
XX lymphoma, breast cancer, prostate cancer or leukemia, or for screening  
XX drug candidates or bioactive agents capable of binding to, or modulating  
XX the activity of, a carcinoma associated protein. The present sequence is  
XX the genomic DNA sequence of the human REL gene which is a carcinoma  
XX associated gene of the invention.  
XX  
SQ Sequence 60940 BP; 16572 A; 11095 C; 10946 G; 18968 T; 0 U; 3359 Other;  
Alignment Scores:  
Pred. No.: 6.34e+04 Length: 60940  
Score: 38.00 Matches: 8  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 1  
Query Match: 86.36% Indels: 0  
DB: 10 Gaps: 0  
US-10-617-443b-2\_COPY\_1\_9 (1-9) x ADE95830 (1-60940)  
QY 1 MechisteleuAlaLeuProAlaSerAla 9  
DB 11468 ATGCATTAGCCCTCCCTGCTTGTCT 11442  
RESULT 9  
AA199682\_19/C.  
Continuation (20 of 45) of AA199682 from base 1900001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682  
WP Fragment Name Begin End  
WP AA199682\_00 1 110000  
WP AA199682\_01 100001 210000  
WP AA199682\_02 200001 310000  
WP AA199682\_03 300001 410000  
WP AA199682\_04 400001 510000  
WP AA199682\_05 500001 610000  
WP AA199682\_06 600001 710000  
WP AA199682\_07 700001 810000  
WP AA199682\_08 800001 910000  
WP AA199682\_09 900001 1010000  
WP AA199682\_10 1000001 1110000  
WP AA199682\_11 1100001 1210000  
WP AA199682\_12 1200001 1310000  
WP AA199682\_13 1300001 1410000  
WP AA199682\_14 1400001 1510000  
WP AA199682\_15 1500001 1610000  
WP AA199682\_16 1600001 1710000  
WP AA199682\_17 1700001 1810000  
WP AA199682\_18 1800001 1910000  
WP AA199682\_19 1900001 2010000  
WP AA199682\_20 2000001 2110000  
WP AA199682\_21 2100001 2210000  
WP AA199682\_22 2200001 2310000  
WP AA199682\_23 2300001 2410000  
WP AA199682\_24 2400001 2510000  
WP AA199682\_25 2500001 2610000  
WP AA199682\_26 2600001 2710000  
WP AA199682\_27 2700001 2810000  
WP AA199682\_28 2800001 2910000

WP AA199682\_29 2900001 3010000  
 WP AA199682\_30 3000001 3110000  
 WP AA199682\_31 3100001 3210000  
 WP AA199682\_32 3200001 3310000  
 WP AA199682\_33 3300001 3410000  
 WP AA199682\_34 3400001 3510000  
 WP AA199682\_35 3500001 3610000  
 WP AA199682\_36 3600001 3710000  
 WP AA199682\_37 3700001 3810000  
 WP AA199682\_38 3800001 3910000  
 WP AA199682\_39 3900001 4010000  
 WP AA199682\_40 4000001 4110000  
 WP AA199682\_41 4100001 4210000  
 WP AA199682\_42 4200001 4310000  
 WP AA199682\_43 4300001 4410000  
 WP AA199682\_44 4400001 4411529

Alignment Scores:  
 Pred. No.: 1.19e+05 Length: 110000  
 Score: 38.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 86.36% Indels: 0  
 DB: Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x AA199682\_19 (1-110000)

OY 1 MethisLeu1aleuPro1aSer1a 9  
 ID AA067901 standard; DNA; 2196 BP.  
 DB 87705 GTTCACTTAGCGATGCGGCTGAGCT 87679

RESULT 10  
 ID AA067901 standard; DNA; 2196 BP.  
 AC AA067901;  
 XX 25-MAR-2003 (revised)  
 DT 08-DEC-1994 (first entry)  
 XX  
 DE Syndecan gene enhancer element.  
 XX  
 DE Syndecan; enhancer; differentiation; tumor; therapy; hair growth; ss.  
 XX  
 OS Mus sp.  
 XX  
 PN WO9412162-A1.  
 XX  
 PD 09-JUN-1994.  
 XX  
 PF 01-DEC-1993; 93WO-F1000514.  
 XX  
 PR 01-DEC-1992; 92US-00988427.  
 XX  
 PA (WAER/) WAERRI A M.  
 PA (ALAN/) ALANEN-KURKI L M.  
 PA (AUVI/) AUVINEN P O V.  
 PA (JAAK/) JAAKKOLA P M.  
 PA (JALK/) JALKANEN M T.  
 PA (LEPP/) LEPPAE S M.  
 PA (MALI/) MALI M S.  
 PA (VIHI/) VIHINEN T A.  
 XX  
 PI Maerri AM, Alanen-Kurki LM, Auvinen POV, Jaakkola PM, Jalkanen MT,  
 PI Leppae SM, Mali MS, Vihtinen TA;  
 DR WPI: 1994-199926/24.  
 XX  
 PT Syndecan stimulation of cellular differentiation - useful for decreasing  
 PT tumour growth used to promote hair growth.  
 XX  
 PS Disclosure; Page 40-41; 65pp; English.  
 XX

CC The mouse syndecan gene enhancer, located 8-10 kb upstream from the  
 CC initiation site, is given in AA067901. Manipulation of the enhancer can  
 CC be used either to slow or prevent tumor growth or to promote  
 CC differentiation of specific cell types, e.g. epidermal cells to promote  
 CC hair formation. The complete mouse syndecan gene and its encoded protein  
 CC are given in AA067902 and AAR55276. (Updated on 25-MAR-2003 to correct  
 CC field.) (Updated on 25-MAR-2003 to correct PA field.) (Updated on 25-MAR-  
 CC 2003 to correct PI field.)  
 XX  
 SQ Sequence 2196 BP; 509 A; 582 C; 536 G; 569 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 2.7e+03 Length: 2196  
 Score: 37.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 87.50% Mismatches: 0  
 Query Match: 84.09% Indels: 0  
 DB: Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x AA067901 (1-2196)

OY 1 MethisLeu1aleuPro1aSer 8  
 ID AA067901 standard; DNA; 2196 BP.  
 DB 1143 ATGCATGTGCACCTCCTGCTCA 1120

RESULT 11  
 ID AA067901 standard; DNA; 2196 BP.  
 AC AA067901;  
 XX 28-SEP-1998 (first entry)  
 DT  
 XX  
 DE Nucleotide sequence of mouse syndecan gene.  
 XX  
 DE Mouse; syndecan gene; syndecan enhancer element; expression vector;  
 XX  
 KW wound healing; ss.  
 XX  
 OS Mus sp.  
 XX  
 PN WO9824921-A1.  
 XX  
 PD 11-JUN-1998.  
 XX  
 PF 02-DEC-1997; 97WO-F1000748.  
 XX  
 PR 02-DEC-1996; 96US-00760534.  
 XX  
 PA (BIOT-) BIOTIE THERAPIES LTD OY.  
 XX  
 PI Jalkanen M, Jaakkola P, Vihtinen T;  
 XX  
 DR WPI: 1998-33338/29.  
 XX  
 PT New syndecan enhancer element - useful in expression vector(s) for  
 PT promoting wound healing.  
 XX  
 PS Disclosure; Page 29-30; 57pp; English.  
 XX  
 CC This is the nucleotide sequence of the mouse syndecan gene, used in the  
 CC method of the invention involving the novel syndecan enhancer element,  
 CC useful in expression vectors. The vector can be used to produce a host  
 CC cell for expressing a structural gene. The expression vector can be used  
 CC to differentially express a gene at a wound site, and promote wound  
 CC healing  
 XX  
 SQ Sequence 2196 BP; 509 A; 582 C; 536 G; 569 T; 0 U; 0 Other;

Low = 276-26127 = 46.3%

Alignment Scores:  
 Pred. No.: 2.7e+03 Length: 2196  
 Score: 37.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 87.50% Mismatches: 0

$\angle C = 90.3^\circ$   
26/27

```

CC healing
XX
SQ Sequence 2196 BP; 509 A; 582 C; 536 G; 569 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,7e+03 Length: 2196
Score: 37.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 84.09% Indels: 0
DB: 2 Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x AAV15947 (1-2196)
OY 1 MethisLeuAlaLeuProAlaSer 8
DB 1143 ATGCATGTGACACTCCCTGCTCA 1120

RESULT 13
ACA60751/c
ID ACA60751 standard; DNA; 2196 BP.
XX
XX ACA60751;
AC
XX
XX 27-JUN-2003 (first entry)
DT
XX
XX Mouse syndecan enhancer element, FGF-inducible Response Element (FIRE)
DE
XX
XX Mouse; db: syndecan enhancer element; cutaneous wound healing; FIRE;
KW tissue regeneration; syndecan regulation; syndecan expression;
KM hair growth; FGF-inducible Response Element.
XX
XX
OS
XX
XX Mus sp.
XX
XX
XX Key location/Qualifiers
FH enhancer /tag= a 150..1499
FT /tag= a /standard name= "Enhancer"
FT enhancer 1331..1614
FT /tag= b /standard name= "Enhancer"
FT /note= "Specifically claimed in claim 1"
FT misc_binding 1407..1446
FT /tag= c /bound motif= "FGF-2 dependent nuclear factor"
FT /note= "DNA binding motif 5"
FT misc_binding 1450..1468
FT /tag= d /bound motif= "FGF-2 dependent nuclear factor"
FT /note= "DNA binding motif 4"
FT misc_binding 1477..1491
FT /tag= e /bound motif= "FGF-2 dependent nuclear factor"
FT /note= "DNA binding motif 3"
FT misc_binding 1497..1511
FT /tag= f /bound motif= "FGF-2 independent nuclear factor"
FT /note= "DNA binding motif 2"
FT misc_binding 1526..1551
FT /tag= g /bound motif= "FGF-2 independent nuclear factor"
FT /note= "DNA binding motif 1"
XX
XX US6492344-B1.
XX
XX 10-DEC-2002.
XX
XX 21-JUN-1999; 99US-00336757.
XX
XX 01-DEC-1993; 93WO-F1000514.
XX 07-MAR-1994; 94US-00206186.
XX
XX (BIOT-) BIOTIE THERAPIES CORP.

```

XX  
PI Jalkanen M, Jaakkola P, Vahinen T;  
XX WPI; 2003-391071/37.  
DR  
XX Promotion of cutaneous wound healing comprises introducing to wound site  
PT recombinant expression vector comprising syndecan enhancer element linked  
PT to promoter which is linked to structural gene encoding growth factor.  
XX  
XX Example 8; Fig 4; 84pp; English.  
XX The invention relates to a method of promoting cutaneous wound healing  
CC which comprises directly introducing to a cutaneous wound site a  
CC recombinant expression vector comprising the syndecan enhancer element,  
CC FGF-inducible Response Element (FIRE), operably linked to a FIRE-  
CC activated promoter which is operably linked to a structural gene encoding  
CC a growth factor that promotes the cutaneous wound healing. The method is  
CC for promoting regeneration of tissue e.g. promotion of hair growth. The  
CC invention enhances syndecan expression via administration of growth  
CC factors. The enhancer element regulates the expression of syndecan. The  
CC present sequence represents the mouse syndecan enhancer element, FGF-  
CC inducible Response Element (FIRE)  
XX  
SQ Sequence 2196 BP; 509 A; 582 C; 536 G; 569 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 2.7e+03 Length: 2196  
Score: 37.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 87.50% Mismatches: 0  
Query Match: 84.09% Indels: 0  
DB: 8 Gaps: 0  
US-10-617-443B-2\_COPY\_1\_9 (1-9) x ACN45004 (1-21562)  
QY 1 MethistenualeuProAlaser 8  
DB 1143 ATGCATGTGGCACTCCCTGCTCA 1120  
RESULT 14  
ACN45004 standard; DNA; 21562 BP.  
XX  
AC ACN45004;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Mouse genomic sequence MCG6875.  
XX  
KM Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.  
XX  
OS Mus musculus.  
XX  
PN MO2003073826-A2.  
XX  
PD 12-SEP-2003.  
XX  
PF 28-FEB-2003; 2003WO-US006235.  
XX  
PR 01-MAR-2002; 2002US-00087192.  
XX  
PA (SAGR-) SAGRES DISCOVERY.  
XX  
PI Morris DW;  
XX  
DR WPI; 2003-328604/31.  
XX  
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
PT comprises a nucleotide sequence.  
XX  
PS Claim 1; SEQ ID NO 1735; ODP; English.  
XX  
CC The present invention relates to novel DNA and protein sequences which

CC are associated with carcinomas. The sequences are useful for: (i) for  
CC screening drug candidates; (ii) for screening of bioactive agent capable  
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
CC evaluating the effect of a candidate carcinoma drug; (v) for treating  
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for creating  
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biologic;  
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
CC carcinoma including lymphoma. The present sequence is one such CA coding  
CC sequence. Note: This patent is an equivalent to basic patent  
CC US2002182586A1, for which no sequence data was published  
XX  
SQ Sequence 21562 BP; 5504 A; 4710 C; 4598 G; 5450 T; 0 U; 1300 Other;  
Alignment Scores:  
Pred. No.: 3.21e+04 Length: 21562  
Score: 37.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 84.09% Indels: 0  
DB: 11 Gaps: 0  
US-10-617-443B-2\_COPY\_1\_9 (1-9) x ACN45004 (1-21562)  
QY 1 MethistenualeuProAlaserA 9  
DB 11652 ATCCATCTTGCGCTGCATCCAGTGCA 11678  
RESULT 15  
AAK67239/C  
ID AAK67239 standard; DNA; 30393 BP.  
XX  
AC AAK67239;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22051.  
XX  
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001354.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0190874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-020515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217486P.  
PR 14-JUL-2000; 2000US-0218280P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.

PR	14-AUG-2000;	2000US-0225266P.
PR	14-AUG-2000;	2000US-0225267P.
PR	14-AUG-2000;	2000US-0225268P.
PR	14-AUG-2000;	2000US-0225270P.
PR	14-AUG-2000;	2000US-0225475P.
PR	14-AUG-2000;	2000US-0225757P.
PR	14-AUG-2000;	2000US-0225758P.
PR	14-AUG-2000;	2000US-0225759P.
PR	18-AUG-2000;	2000US-0226279P.
PR	22-AUG-2000;	2000US-0226681P.
PR	22-AUG-2000;	2000US-0227182P.
PR	23-AUG-2000;	2000US-0227182P.
PR	23-AUG-2000;	2000US-0227109P.
PR	30-AUG-2000;	2000US-0228928P.
PR	01-SEP-2000;	2000US-0229287P.
PR	01-SEP-2000;	2000US-0229343P.
PR	01-SEP-2000;	2000US-0229344P.
PR	01-SEP-2000;	2000US-0229345P.
PR	05-SEP-2000;	2000US-0229509P.
PR	05-SEP-2000;	2000US-0229513P.
PR	06-SEP-2000;	2000US-0230437P.
PR	06-SEP-2000;	2000US-0230438P.
PR	08-SEP-2000;	2000US-0231243P.
PR	08-SEP-2000;	2000US-0231243P.
PR	08-SEP-2000;	2000US-0231244P.
PR	08-SEP-2000;	2000US-0231413P.
PR	08-SEP-2000;	2000US-0231414P.
PR	08-SEP-2000;	2000US-0232080P.
PR	08-SEP-2000;	2000US-0232081P.
PR	12-SEP-2000;	2000US-0231968P.
PR	14-SEP-2000;	2000US-0233397P.
PR	14-SEP-2000;	2000US-0233398P.
PR	14-SEP-2000;	2000US-0233399P.
PR	14-SEP-2000;	2000US-0233400P.
PR	14-SEP-2000;	2000US-0233401P.
PR	14-SEP-2000;	2000US-0233064P.
PR	14-SEP-2000;	2000US-0233064P.
PR	21-SEP-2000;	2000US-0233065P.
PR	21-SEP-2000;	2000US-0234223P.
PR	21-SEP-2000;	2000US-0234274P.
PR	25-SEP-2000;	2000US-0234997P.
PR	25-SEP-2000;	2000US-0234998P.
PR	26-SEP-2000;	2000US-0235484P.
PR	27-SEP-2000;	2000US-0235834P.
PR	27-SEP-2000;	2000US-0235835P.
PR	29-SEP-2000;	2000US-0236327P.
PR	29-SEP-2000;	2000US-0236367P.
PR	29-SEP-2000;	2000US-0236368P.
PR	29-SEP-2000;	2000US-0236369P.
PR	29-SEP-2000;	2000US-0236370P.
PR	02-OCT-2000;	2000US-0237037P.
PR	02-OCT-2000;	2000US-0237038P.
PR	02-OCT-2000;	2000US-0237039P.
PR	02-OCT-2000;	2000US-0237040P.
PR	13-OCT-2000;	2000US-0239335P.
PR	13-OCT-2000;	2000US-0239337P.
PR	20-OCT-2000;	2000US-0240960P.
PR	20-OCT-2000;	2000US-0241221P.
PR	20-OCT-2000;	2000US-0241785P.
PR	20-OCT-2000;	2000US-0241786P.
PR	20-OCT-2000;	2000US-0241787P.
PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225266P.
PR	14-AUG-2000;	2000US-0225267P.
PR	14-AUG-2000;	2000US-0225268P.
PR	14-AUG-2000;	2000US-0225270P.
PR	14-AUG-2000;	2000US-0225475P.
PR	14-AUG-2000;	2000US-0225757P.
PR	14-AUG-2000;	2000US-0225758P.
PR	14-AUG-2000;	2000US-0225759P.
PR	18-AUG-2000;	2000US-0226279P.
PR	22-AUG-2000;	2000US-0226681P.
PR	22-AUG-2000;	2000US-0227182P.
PR	23-AUG-2000;	2000US-0227182P.
PR	30-AUG-2000;	2000US-0228928P.
PR	01-SEP-2000;	2000US-0229287P.
PR	01-SEP-2000;	2000US-0229343P.
PR	01-SEP-2000;	2000US-0229344P.
PR	01-SEP-2000;	2000US-0229345P.
PR	05-SEP-2000;	2000US-0229509P.
PR	05-SEP-2000;	2000US-0229513P.
PR	06-SEP-2000;	2000US-0230437P.
PR	06-SEP-2000;	2000US-0230

Alignment Scores:  
 Pred. No.: 4,666+04 Length: 30393  
 Score: 37.00 Matches: 7  
 Percent Similarity: 88.89% Conservative: 1  
 Best Local Similarity: 77.78% Mismatches: 1  
 Query Match: 84.09% Indels: 0  
 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x AAK67239 (1-30393)

Qy 1 MetHisLeuAlaLeuProAlaSerAla 9  
 Db 16409 ATGCACCTCACCCCTCCCTAGTAGCGCT 16383

Search completed: June 29, 2005, 21:40:29  
 Job time : 50.2682 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 29, 2005, 21:23:47 ; Search time 28.1808 Seconds  
(without alignments)  
1996.208 Million cell updates/sec

Title: US-10-617-443B-2\_COPY\_1\_9  
Perfect score: 44  
Sequence: 1 MHLALPASA 9

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 6067389 seqs, 3125258755 residues  
Total number of hits satisfying chosen parameters: 12134778

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.epool/US10617443/rnmb.29062005.135933.26126/app\_query.fasta.1.718  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=trmb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=tblomsum62  
-TRANS=human40.cst -LIST=45 -DOCLIGN=200 -THR SCORE=oct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MATEL=200000000 -USER=US10617443@cgn\_1\_1582@rnmb.29062005.135933.26126  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEDBTRY -NEG SCORE=0 -WAIT -DSBLOCL=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database: Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PTCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PTCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:\*  
24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	3006	17 US-10-189-099A-1	Sequence 1, Appl
2	44	100.0	17	US-10-617-443B-1	Sequence 1, Appl
3	39	88.6	754	9 US-09-910-943-654	Sequence 654, App
4	39	88.6	963	17 US-10-282-122A-13408	Sequence 13408, A
5	39	88.6	2157	17 US-10-282-122A-13840	Sequence 13840, A
6	38	86.4	1500	19 US-10-437-963-44767	Sequence 44767, A
7	38	86.4	1581	18 US-10-424-599-10035	Sequence 10035, A
8	38	86.4	60940	18 US-10-052-482-88	Sequence 88, Appl
9	37	84.1	582	13 US-10-027-633-270357	Sequence 270357, A
10	37	84.1	582	17 US-10-027-633-270357	Sequence 270357, A
11	37	84.1	4513	19 US-10-437-963-44767	Sequence 44767, A
12	37	84.1	21562	13 US-10-087-193-1735	Sequence 1735, Ap
13	37	84.1	24754	19 US-10-322-696-55	Sequence 55, Appl
14	36	81.8	34	17 US-10-189-099A-5	Sequence 5, Appl
15	36	81.8	34	21 US-10-617-443B-5	Sequence 5, Appl
16	36	81.8	281	20 US-10-425-115-88970	Sequence 88970, A
17	36	81.8	557	19 US-10-437-963-16475	Sequence 16475, A
18	36	81.8	582	13 US-10-027-633-15114	Sequence 15114, A
19	36	81.8	582	13 US-10-027-633-15115	Sequence 15115, A
20	36	81.8	582	13 US-10-027-633-15116	Sequence 15116, A
21	36	81.8	582	17 US-10-027-633-15114	Sequence 15114, A
22	36	81.8	582	17 US-10-027-633-15115	Sequence 15115, A
23	36	81.8	582	17 US-10-027-633-15116	Sequence 15116, A
24	36	81.8	756	17 US-10-282-122A-10678	Sequence 10678, A
25	36	81.8	757	17 US-10-282-122A-9829	Sequence 9829, Ap
26	36	81.8	778	20 US-10-425-115-166607	Sequence 166607, A
27	36	81.8	953	13 US-10-027-633-121792	Sequence 121792, A
28	36	81.8	1138	17 US-10-027-633-121792	Sequence 121792, A
29	36	81.8	1138	13 US-10-027-633-118836	Sequence 118836, A
30	36	81.8	1138	17 US-10-027-633-118836	Sequence 118836, A
31	36	81.8	1854	15 US-10-037-270-992	Sequence 992, App
32	36	81.8	1854	17 US-10-117-722-992	Sequence 992, App
33	36	81.8	1922	20 US-10-425-115-117912	Sequence 117912, A
34	36	81.8	1967	19 US-10-437-963-87894	Sequence 87894, A
35	36	81.8	2160	17 US-10-364-888-7	Sequence 7, Appl
36	36	81.8	2203	17 US-10-309-292-113	Sequence 113, App
37	36	81.8	2250	16 US-10-283-976-22	Sequence 22, Appl
38	36	81.8	2250	21 US-10-847-976-21	Sequence 21, Appl
39	36	81.8	2491	17 US-10-094-749-803	Sequence 803, App
40	36	81.8	2566	18 US-10-336-603A-1	Sequence 1, Appl
41	36	81.8	2573	19 US-10-767-701-14417	Sequence 14417, A
42	36	81.8	2629	17 US-10-309-292-115	Sequence 115, App
43	36	81.8	2707	17 US-10-291-117-6	Sequence 6, Appl
44	36	81.8	2707	18 US-10-221-278-6	Sequence 6, Appl
45	36	81.8	2714	17 US-10-094-749-1018	Sequence 1018, Ap

ALIGNMENTS

RESULT 1  
US-10-189-099A-1  
; Sequence 1, Application US/10189099A  
; Publication No. US20040005659A1  
; GENERAL INFORMATION:  
; APPLICANT: Darrow, Andrew L  
; APPLICANT: Qi, Jian-shen  
; APPLICANT: Chen, Gailin  
; APPLICANT: Andrade-Gordon, Patricia  
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof  
; FILE REFERENCE: CRT-1644  
; CURRENT APPLICATION NUMBER: US/10/189, 099A  
; CURRENT FILING DATE: 2002-12-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1

```
; LENGTH: 3006
; TYPE: CDNA
; ORGANISM: Homo sapiens
US-10-189-099A-1

Alignment Scores:
Pred. No.: 54.2      Length: 3006
Score: 44.00      Matches: 9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 17      Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-189-099A-1 (1-3006)
QY 1 MethiLeuAlaleuProAlaSerAla 9
DB 1011 ATGCACCTGGCCCTCCGCCGACGCA 1037

RESULT 2
US-10-617-443B-1
; Sequence 1, Application US/10617443B
; Publication No. US2005001977A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Caillin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/617,443B
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3006
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-617-443B-1

Alignment Scores:
Pred. No.: 54.2      Length: 3006
Score: 44.00      Matches: 9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 21      Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-617-443B-1 (1-3006)
QY 1 MethiLeuAlaleuProAlaSerAla 9
DB 1011 ATGCACCTGGCCCTCCGCCGACGCA 1037

RESULT 3
US-09-910-943-654
; Sequence 654, Application US/09910943
; Patent No. US20020081610A1
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Bivanolou, Ali
; APPLICANT: Altman, Curtis
; TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
; FILE REFERENCE: 7529/IG148US1
; CURRENT APPLICATION NUMBER: US/09/910,943
; NUMBER OF SEQ ID NOS: 742
; CURRENT FILING DATE: 2001-07-23
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 654
; LENGTH: 754
; TYPE: DNA
; ORGANISM: Xenopus laevis
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (1)..(754)
; OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-654

Alignment Scores:
Pred. No.: 147      Length: 754
Score: 39.00      Matches: 7
Percent Similarity: 100.00%      Conservative: 2
Best Local Similarity: 77.78%      Mismatches: 0
Query Match: 88.64%      Indels: 0
DB: 9      Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-910-943-654 (1-754)
QY 1 MethiLeuAlaleuProAlaSerAla 9
DB 327 TTACATATGCTCTGCCGCTCTGCT 353

RESULT 4
US-10-282-122A-13408
; Sequence 13408, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13408
; LENGTH: 963
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-282-122A-13408

Alignment Scores:
Pred. No.: 189      Length: 963
Score: 39.00      Matches: 8
Percent Similarity: 88.89%      Conservative: 1
Best Local Similarity: 88.89%      Mismatches: 1
```



Query Match: 88.64% Indels: 0  
DB: 17 Gaps: 0

US-10-617-443b-2\_COPY\_1\_9 (1-9) x US-10-282-122A-13840 (1-963)

QY 1 MethiSleuA1aleuProA1aSerA1a 9  
DB ATGCATCTGGCCAGCGCGCTCGCGC 765

RESULT 5

US-10-282-122A-13840  
Sequence 13840, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITTA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 13840  
LENGTH: 2157  
TYPE: DNA  
ORGANISM: Burkholderia mallei  
US-10-282-122A-13840

Alignment Scores:  
Pred. No.: 429 Length: 2157  
Score: 39.00 Matches: 8  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 1  
Query Match: 88.64% Indels: 0  
Gaps: 0

US-10-617-443b-2\_COPY\_1\_9 (1-9) x US-10-282-122A-13840 (1-2157)

QY 1 MethiSleuA1aleuProA1aSerA1a 9  
DB 1933 ATGCATCTGGCCAGCGCGCTCGCGC 1959

RESULT 6  
US-10-437-963-43712  
Sequence 43712, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 43712  
LENGTH: 1500  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_46842C.1  
US-10-437-963-43712

Alignment Scores:  
Pred. No.: 479 Length: 1500  
Score: 38.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 86.36% Indels: 0  
Gaps: 0

US-10-617-443b-2\_COPY\_1\_9 (1-9) x US-10-437-963-43712 (1-1500)

QY 1 MethiSleuA1aleuProA1aSerA1a 9  
DB 201 CTCACCTGCTCTCCCGCGCGCA 227

RESULT 7  
US-10-424-599-10035  
Sequence 10035, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 10035  
LENGTH: 1581  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_109071C.1  
US-10-424-599-10035

Alignment Scores:  
Pred. No.: 506 Length: 1581  
Score: 38.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 86.36% Indels: 0  
Gaps: 0

US-10-617-443b-2\_COPY\_1\_9 (1-9) x US-10-424-599-10035 (1-1581)

Thu Jun 30 08:48:38 2005

us-10-617-443b-2\_copy\_1\_9.rmpb

Page 4

```
OY      1 MethisLeuAlaleuProAlaSerAla 9
DB      1155 CTGCATCTAGCGCTCCACAGCAGCTGCA 1181

RESULT 8
US-10-052-482-88/c
; Sequence 88, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 09/798,586
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 88
; LENGTH: 60940
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5047)..(7943)
; OTHER INFORMATION: "n" at positions 5047 to 7943 can be any base
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (35382)..(35843)
; OTHER INFORMATION: "n" at positions 35382 to 35843 can be any base
US-10-052-482-88

Alignment Scores:
Pred. No.:      2.09e+04      Length:      60940
Score:          38.00         Matches:      8
Percent Similarity: 88.89%     Conservative: 0
Best Local Similarity: 88.89%     Mismatches:  1
Query Match:    86.36%         Indels:      0
DB:             18            Gaps:         0

US-10-617-443b-2_COPY_1_9 (1-9) x US-10-052-482-88 (1-60940)

OY      1 MethisLeuAlaleuProAlaSerAla 9
DB      11468 ATGCATTAGCCCTCCCTGCTTGTCT 11442

RESULT 9
US-10-027-632-270357/c
; Sequence 270357, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
```

```
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 270357
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-270357

Alignment Scores:
Pred. No.:      295          Length:      582
Score:          37.00         Matches:      7
Percent Similarity: 88.89%     Conservative:  1
Best Local Similarity: 77.78%     Mismatches:  1
Query Match:    84.09%         Indels:      0
DB:             13            Gaps:         0

US-10-617-443b-2_COPY_1_9 (1-9) x US-10-027-632-270357 (1-582)

OY      1 MethisLeuAlaleuProAlaSerAla 9
DB      495 CTACACCTGGTACTGCTGCTCGCT 469

RESULT 10
US-10-027-632-270357/c
; Sequence 270357, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 270357
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-270357

Alignment Scores:
Pred. No.:      295          Length:      582
Score:          37.00         Matches:      7
Percent Similarity: 88.89%     Conservative:  1
Best Local Similarity: 77.78%     Mismatches:  1
Query Match:    84.09%         Indels:      0
DB:             17            Gaps:         0

US-10-617-443b-2_COPY_1_9 (1-9) x US-10-027-632-270357 (1-582)

OY      1 MethisLeuAlaleuProAlaSerAla 9
DB      495 CTACACCTGGTACTGCTGCTCGCT 469

RESULT 11
US-10-437-963-44767
```

Thu Jun 30 08:48:38 2005

```

; Sequence 44767, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 44767
; LENGTH: 4513
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_47796C.1
; US-10-437-963-44767

```

```

Alignment Scores:
Pred. No.: 2,38e+03 Length: 4513
Score: 37.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 84.09% Indels: 0
DB: 19 Gaps: 0

```

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-10-437-963-44767 (1-4513)

```

Qy 2 HisleuAlaleuProAlaSerAla 9
Db 4231 CACCTACGATGCTGCTCGCGC 4254

```

```

RESULT 12
US-10-087-192-1735
; Sequence 1735, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1735
; LENGTH: 21562
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(21562)
; OTHER INFORMATION: n = A,T,C or G
; US-10-087-192-1735

```

```

Alignment Scores:
Pred. No.: 1.17e+04 Length: 21562
Score: 37.00 Matches: 7
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 77.78% Mismatches: 0
Query Match: 84.09% Indels: 0

```

```

DB: 13 Gaps: 0
US-10-617-443B-2_COPY_1_9 (1-9) x US-10-087-192-1735 (1-21562)
Qy 1 MethisluAlaleuProAlaSerAla 9
Db 11652 ATCCATCTTCGCTGCGCATCCAGTCGA 11678

```

```

RESULT 13
US-10-322-696-55
; Sequence 55, Application US/10322696
; Publication No. US20040166490A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Malandro, Marc
; TITLE OF INVENTION: NOVEL THERAPEUTIC TARGETS IN CANCER
; FILE REFERENCE: 529452001200
; CURRENT APPLICATION NUMBER: US/10/322,696
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 55
; LENGTH: 247544
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(247544)
; OTHER INFORMATION: n = A,T,C or G
; US-10-322-696-55

```

```

Alignment Scores:
Pred. No.: 1.4e+05 Length: 247544
Score: 37.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 84.09% Indels: 0
DB: 19 Gaps: 0

```

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-10-322-696-55 (1-247544)

```

Qy 1 MethisluAlaleuProAlaSerAla 9
Db 23468 CTCATCTTCCTTTCACGCTTCTGCA 23494

```

```

RESULT 14
US-10-189-099A-5
; Sequence 5, Application US/10189099A
; Publication No. US20040005659A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Callin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/189,099A
; CURRENT FILING DATE: 2002-12-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 34
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
; US-10-189-099A-5

```

```

Alignment Scores:
Pred. No.: 26.5 Length: 34
Score: 36.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 81.82% Indels: 0  
DB: 17 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-10-189-099A-5 (1-34)

QY 1 MethisLeuAlaLeuProAla 7  
|||||  
DB 14 ATGCACCTGCGCCCTTCCCGCC 34

RESULT 15

US-10-617-443B-5  
; Sequence 5, Application US/10617443B  
; Publication No. US2005001977A1

; GENERAL INFORMATION:

; APPLICANT: Darrow, Andrew L

; APPLICANT: Qi, Jian-shen

; APPLICANT: Chen, Caillin

; APPLICANT: Andrade-Gordon, Patricia

; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof

; FILE REFERENCE: ORT-1644

; CURRENT APPLICATION NUMBER: US/10/617,443B

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5

; LENGTH: 34

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: primer

US-10-617-443B-5

Alignment Scores:  
Pred. No.: 26.5 Length: 34  
Score: 36.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 81.82% Indels: 0  
DB: 21 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-10-617-443B-5 (1-34)

QY 1 MethisLeuAlaLeuProAla 7  
|||||  
DB 14 ATGCACCTGCGCCCTTCCCGCC 34

Search completed: June 30, 2005, 01:51:25  
Job time : 52.1808 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 29, 2005, 21:11:10 ; Search time 8.02915 Seconds

(without alignments)  
1834.127 Million cell updates/sec

Title: US-10-617-443B-2\_COPY\_1\_9  
Perfect score: 44  
Sequence: 1 MHLAPASA 9

Scoring table:  
BLOSUM62  
Xgapop 10.0, Ygapext 0.5  
Xgapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.epool/US10617443/runat\_29062005\_135932\_26096/app\_query.fasta\_1.718  
-DB=Issued Patents NA -QPMT=fastcap -SUFFIX=rml -MINMATCH=0.1 -LOOFCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPRT=pco -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10617443.qcgn\_1\_1\_93@runat\_29062005\_135932\_26096 -NCPU=3  
-NO MAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGCLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	38	86.4	4411529	3	US-09-103-840A-1
C 2	37	84.1	601	4	US-09-949-016-40948
C 3	37	84.1	601	4	US-09-949-016-40949
C 4	37	84.1	601	4	US-09-949-016-40950
C 5	37	84.1	601	4	US-09-949-016-40951
C 6	37	84.1	601	4	US-09-949-016-40952
C 7	37	84.1	601	4	US-09-949-016-107301
C 8	37	84.1	2196	4	US-08-472-217-3
C 9	37	84.1	2196	3	US-08-760-534A-3
C 10	37	84.1	2196	4	US-09-336-757-3
C 11	37	84.1	35688	4	US-09-949-016-16873
C 12	37	84.1	150780	4	US-09-949-016-14711

C 13	37	84.1	462589	4	US-09-949-016-12900	Sequence 12900, A
C 14	37	84.1	476044	4	US-09-949-016-12412	Sequence 12412, A
C 15	36	81.8	1854	4	US-09-620-312D-992	Sequence 992, App
C 16	36	81.8	5692	4	US-09-902-540-783	Sequence 783, App
C 17	36	81.8	9143	2	US-08-639-857-32	Sequence 32, App1
C 18	36	81.8	9143	3	US-08-469-260A-390	Sequence 390, App
C 19	36	81.8	9143	3	US-08-469-260A-393	Sequence 393, App
C 20	36	81.8	9143	4	US-08-469-260A-393	Sequence 393, App
C 21	36	81.8	9143	4	US-08-469-260A-393	Sequence 393, App
C 22	36	81.8	9143	4	US-08-467-344A-390	Sequence 390, App
C 23	36	81.8	9143	4	US-08-467-344A-393	Sequence 393, App
C 24	36	81.8	9143	4	US-08-424-550B-390	Sequence 390, App
C 25	36	81.8	9143	4	US-08-424-550B-393	Sequence 393, App
C 26	36	81.8	18200	4	US-09-949-016-15661	Sequence 15661, A
C 27	36	81.8	18200	4	US-09-949-016-15661	Sequence 15661, A
C 28	36	81.8	19719	4	US-09-949-016-15662	Sequence 15662, A
C 29	36	81.8	19719	4	US-09-949-016-15663	Sequence 15663, A
C 30	36	81.8	150032	4	US-09-949-016-14321	Sequence 14321, A
C 31	35	79.5	601	4	US-09-949-016-76709	Sequence 76709, A
C 32	35	79.5	601	4	US-09-949-016-76710	Sequence 76710, A
C 33	35	79.5	601	4	US-09-949-016-145816	Sequence 145816, A
C 34	35	79.5	601	4	US-09-949-016-146084	Sequence 146084, A
C 35	35	79.5	601	4	US-09-949-016-146352	Sequence 146352, A
C 36	35	79.5	601	4	US-09-949-016-160820	Sequence 160820, A
C 37	35	79.5	601	4	US-09-949-016-160821	Sequence 160821, A
C 38	35	79.5	601	4	US-09-949-016-162623	Sequence 162623, A
C 39	35	79.5	1192	1	US-08-340-539A-3	Sequence 3, App1
C 40	35	79.5	1192	2	US-08-461-592B-3	Sequence 3, App1
C 41	35	79.5	1362	4	US-09-270-767-12819	Sequence 12819, A
C 42	35	79.5	1469	4	US-09-865-879-12	Sequence 12, App1
C 43	35	79.5	2188	4	US-09-620-312D-409	Sequence 409, App
C 44	35	79.5	2275	4	US-09-620-312D-410	Sequence 410, App
C 45	35	79.5	3224	3	US-08-965-729A-2	Sequence 2, App1

## ALIGNMENTS

RESULT 1  
US-09-103-840A-1/C  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FRISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R. M.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103.840A  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

## Alignment Scores:

Pred. No.: 1.16e+06  
Score: 38.00  
Percent Similarity: 100.00%  
Best Local Similarity: 77.78%  
Query Match: 86.36%  
DB: 3  
Caps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-09-103-840A-1 (1-4411529)

Oy 1 Methisulalaleuproalaserala 9  
Db 1987705 GTTCACTTAGCGATCGCGGCTCAGCT 1987679

```
RESULT 2
US-09-949-016-40948
; Sequence 40948, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40948
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-40948

Alignment Scores:
Pred. No.: 164
Score: 37.00
Percent Similarity: 100.00%
Best Local Similarity: 87.50%
Query Match: 84.09%
DB: 4
Gaps: 0

US-10-617-443b-2_COPY_1_9 (1-9) x US-09-949-016-40948 (1-601)
QY 1 MechIsleuAlaLeuProAlaSer 8
Db 339 ATGCATCTAGCTTTGCCAAGTTCT 362

RESULT 3
US-09-949-016-40949
; Sequence 40949, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40949
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-40949

Alignment Scores:
Pred. No.: 164
Score: 37.00
Percent Similarity: 100.00%
Best Local Similarity: 87.50%
Query Match: 84.09%
DB: 4
Gaps: 0
```

```
US-10-617-443b-2_COPY_1_9 (1-9) x US-09-949-016-40949 (1-601)
QY 1 MechIsleuAlaLeuProAlaSer 8
Db 340 ATGCATCTAGCTTTGCCAAGTTCT 363

RESULT 4
US-09-949-016-40950
; Sequence 40950, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40950
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-40950

Alignment Scores:
Pred. No.: 164
Score: 37.00
Percent Similarity: 100.00%
Best Local Similarity: 87.50%
Query Match: 84.09%
DB: 4
Gaps: 0

US-10-617-443b-2_COPY_1_9 (1-9) x US-09-949-016-40950 (1-601)
QY 1 MechIsleuAlaLeuProAlaSer 8
Db 344 ATGCATCTAGCTTTGCCAAGTTCT 367

RESULT 5
US-09-949-016-40951
; Sequence 40951, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40951
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-40951

Alignment Scores:
Pred. No.: 164
Score: 37.00
```

Percent Similarity: 100.00%  
Best Local Similarity: 87.50%  
Query Match: 84.09%  
DB: 4

Conservative: 1  
Matches: 0  
Indels: 0  
Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-09-949-016-40951 (1-601)

QY 1 MethislaeuAlaepProAlaSer 8  
DB 455 ATGCATCTAGCTTGCCAGTTCT 478

RESULT 6

US-09-949-016-40952  
Sequence 40952, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C0001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 40952  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-40952

# Alignment Scores:

Pred. No.: 164  
Score: 37.00  
Percent Similarity: 100.00%  
Best Local Similarity: 87.50%  
Query Match: 84.09%  
DB: 4

Length: 601  
Matches: 7  
Conservative: 1  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-09-949-016-40952 (1-601)

QY 1 MethislaeuAlaepProAlaSer 8  
DB 479 ATGCATCTAGCTTGCCAGTTCT 502

RESULT 7

US-09-949-016-107301  
Sequence 107301, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C0001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 107301  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human

US-09-949-016-107301

# Alignment Scores:

Pred. No.: 164  
Score: 37.00  
Percent Similarity: 100.00%  
Best Local Similarity: 87.50%  
Query Match: 84.09%  
DB: 4

Length: 601  
Matches: 7  
Conservative: 1  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-09-949-016-107301 (1-601)

QY 1 MethislaeuAlaepProAlaSer 8  
DB 314 TTGCATTTGGCAGCTTCTCTCTCT 337

RESULT 8

US-08-472-217-3/C  
Sequence 3, Application US/08472217  
Patent No. 5726058  
GENERAL INFORMATION:

APPLICANT: Alanen-Kurki, Leena  
APPLICANT: Auvinen, Petri  
APPLICANT: Jaakkola, Panu  
APPLICANT: Jalkanen, Markku  
APPLICANT: Lepp, Sirpa  
APPLICANT: Mali, Markku  
APPLICANT: Vihinen, Tapani  
TITLE OF INVENTION: SynDecan Stimulation Of Cellular  
TITLE OF INVENTION: Differentiation  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,217  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/206,186  
FILING DATE: 07-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/988,427  
FILING DATE: 01-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Cimbal, Michele A.  
REGISTRATION NUMBER: 33,851  
REFERENCE/DOCKET NUMBER: 1102.0050003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2196 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-472-217-3

# Alignment Scores:

Pred. No.: 666 Length: 2196  
 Score: 37.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 87.50% Mismatches: 0  
 Query Match: 84.09% Indels: 0  
 DB: 1 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-08-472-217-3 (1-2196)

QY 1 MethiLeuAlaLeuProAlaSer 8

DB 1143 ATGATGTGGCACTCCCTGCTCA 1120

RESULT 9

US-08-760-534A-3/C

Sequence 3, Application US/08760534A

Patent No. 6017727

GENERAL INFORMATION:

APPLICANT: JALKANEN, MARKKU

APPLICANT: JAAKKOLA, PANU

APPLICANT: VIHINEN, TAPANI

TITLE OF INVENTION: SYNDSCAN ENHANCER ELEMENT AND SYNDSCAN

TITLE OF INVENTION: STIMULATION OF CELLULAR DIFFERENTIATION

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESSES:

ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 NEW YORK AVENUE, SUITE 600

CITY: WASHINGTON

STATE: DC

COUNTRY: US

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/760.534A

FILING DATE: 02-DEC-1996

CLASSIFICATION: 435

APPLICATION DATA:

PRIOR APPLICATION NUMBER: US 08/206,186

FILING DATE: 07-MAR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FI93/00514

DB 1143 ATGATGTGGCACTCCCTGCTCA 1120

RESULT 10

US-09-336-757-3/C

Sequence 3, Application US/09336757

Patent No. 6492344

GENERAL INFORMATION:

APPLICANT: JALKANEN, MARKKU

APPLICANT: JAAKKOLA, PANU

APPLICANT: VIHINEN, TAPANI

TITLE OF INVENTION: SYNDSCAN ENHANCER ELEMENT AND SYNDSCAN

TITLE OF INVENTION: STIMULATION OF CELLULAR DIFFERENTIATION

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESSES:

ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 NEW YORK AVENUE, SUITE 600

CITY: WASHINGTON

STATE: DC

COUNTRY: US

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/336,757

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/760.534

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FI93/00514

FILING DATE: 01-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: CIMBALA, MICHELE A.

DB 1143 ATGATGTGGCACTCCCTGCTCA 1120

RESULT 11

US-09-949-016-16873

Sequence 16873, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESSES:

ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 NEW YORK AVENUE, SUITE 600

CITY: WASHINGTON

STATE: DC

COUNTRY: US

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/336,757

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/760.534

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FI93/00514

FILING DATE: 01-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: CIMBALA, MICHELE A.

REGISTRATION NUMBER: 33,851

REFERENCE/DOCKET NUMBER: 1708.0050004/MAC

TELEPHONE: (202) 371-2600

DB 1143 ATGATGTGGCACTCCCTGCTCA 1120

ALIGNMENT SCORES:

Pred. No.: 666 Length: 2196

Score: 37.00 Matches: 7

Percent Similarity: 100.00% Conservative: 1

Best Local Similarity: 87.50% Mismatches: 0

Query Match: 84.09% Indels: 0

DB: 4 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-09-336-757-3 (1-2196)

QY 1 MethiLeuAlaLeuProAlaSer 8

DB 1143 ATGATGTGGCACTCCCTGCTCA 1120

RESULT 11

US-09-949-016-16873

Sequence 16873, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESSES:

ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 NEW YORK AVENUE, SUITE 600

CITY: WASHINGTON

STATE: DC

COUNTRY: US

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/336,757

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:



```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16873
; LENGTH: 35688
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16873

Alignment Scores:
Pred. No.: 1.36e+04 Length: 35688
Score: 37.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 84.09% Indels: 0
Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-16873 (1-35688)

Qy 1 MethisLeuAlaLeuProAlaSer 8
Db 15904 ATGCACCTCACCTTCCTCCAGTACGCGCT 15930

RESULT 12
US-09-949-016-14711
; Sequence 14711, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14711
; LENGTH: 150780
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14711

Alignment Scores:
Pred. No.: 6.4e+04 Length: 150780
Score: 37.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 84.09% Indels: 0
Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-14711 (1-150780)

Qy 1 MethisLeuAlaLeuProAlaSer 8
Db 42768 TTGCATTGGCATTCTCTGCTTCT 42791

RESULT 13
US-09-949-016-12900/c

```

```

; Sequence 12900, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12900
; LENGTH: 462589
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12900

Alignment Scores:
Pred. No.: 2.09e+05 Length: 462589
Score: 37.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 84.09% Indels: 0
Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-12900 (1-462589)

Qy 1 MethisLeuAlaLeuProAlaSer 8
Db 431394 ATGCATCTAGCTTGGCCAGTTCT 431371

RESULT 14
US-09-949-016-12412/c
; Sequence 12412, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12412
; LENGTH: 476044
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12412

Alignment Scores:
Pred. No.: 2.15e+05 Length: 476044
Score: 37.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 84.09% Indels: 0
Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-12412 (1-476044)

Qy 1 MethisLeuAlaLeuProAlaSer 8

```

DB 444850 ATGCATCTAGCTTGCACAGTTCT 444827

RESULT 15  
US-09-620-312D-992/C

/ Sequence 992, Application US/09620312D  
/ Patent No. 6569662  
/ GENERAL INFORMATION:  
/ APPLICANT: Tang, Y. Tom  
/ APPLICANT: Liu, Chenghua  
/ APPLICANT: Beundt, Vinod  
/ APPLICANT: Zhang, Jie  
/ APPLICANT: Ren, Feiyan  
/ APPLICANT: Chen, Rui-hong  
/ APPLICANT: Zhao, Qing A.  
/ APPLICANT: Mehrman, Tom  
/ APPLICANT: Xue, Aidong J.  
/ APPLICANT: Yang, Yonghong  
/ APPLICANT: Wang, Jian-Rui  
/ APPLICANT: Zhou, Ping  
/ APPLICANT: Ma, Yungqing  
/ APPLICANT: Wang, Duntui  
/ APPLICANT: Wang, Zhiwei  
/ APPLICANT: John Tillinghast  
/ APPLICANT: Drmanac, Radoje T.  
/ TITLE OF INVENTION: Polypeptides  
/ FILE REFERENCE: 784CIP28  
/ CURRENT APPLICATION NUMBER: US/09/620,312D  
/ CURRENT FILING DATE: 2000-07-19  
/ PRIOR APPLICATION NUMBER: 09/552,317  
/ PRIOR FILING DATE: 2000-04-25  
/ PRIOR APPLICATION NUMBER: 09/488,725  
/ PRIOR FILING DATE: 2000-01-21  
/ NUMBER OF SEQ ID NOS: 1105  
/ SOFTWARE: pc\_fl\_genes Version 1.0  
/ SEQ ID NO 992  
/ LENGTH: 1854  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: CDS  
/ LOCATION: (559)..(1710)  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (1)..(1854)  
/ OTHER INFORMATION: n = a,t,c or g  
US-09-620-312D-992

Alignment Scores:  
Pred. No.: 874 Length: 1854  
Score: 36.00 Matches: 7  
Percent Similarity: 88.89% Conservative: 1  
Best Local Similarity: 77.78% Mismatches: 1  
Query Match: 81.82% Indels: 0  
DB: 4 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x US-09-620-312D-992 (1-1854)

QY 1 MetH1sleuAlaIeuPrcAlaSerAla 9  
DB 610 ATGCACCTGCGAGCTGCTGCAACAGCT 584

Search completed: June 30, 2005, 01:41:07  
Job time : 496.029 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 29, 2005, 21:07:15 ; Search time 155.519 Seconds  
(without alignments)  
2202.808 Million cell updates/sec

Title: US-10-617-443B-2\_COPY\_1\_9

Perfect score: 44

Sequence: 1 MHUAPASA 9

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Xgapop 10.0, Xgapext 0.5  
Xgapop 6.0, Xgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10617443/runat\_29062005\_135932\_26076/apd\_query.fasta\_1.718  
-DB=EST -Qfmt=fastap -SUFPr=rc -MINMATCH=0.1 -LOOPCL=0 -LOOEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=00 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US10617443 @CGN 1.1 4385 @runat\_29062005\_135932\_26076 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGESUBSTRY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: EST:

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hlc.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_ges1.\*  
9: gb\_ges2.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	41	93.2	150	4	BJ457822 BJ457822
2	41	93.2	181	4	BJ474645 BJ474645
3	41	93.2	202	4	BJ450407 BJ450407
4	41	93.2	234	4	CD662805 CD662805
5	41	93.2	308	1	AV912195 AV912195
6	41	93.2	314	6	CD662622 UCRHV18_0
7	41	93.2	329	1	AV909564 AV909564
8	41	93.2	332	1	AV911456 AV911456
9	41	93.2	419	1	AV912337 AV912337

C	10	41	93.2	427	1	AV910159	AV910159
C	11	41	93.2	440	4	BJ461882	BJ461882
C	12	41	93.2	446	4	BJ455746	BJ455746
C	13	41	93.2	469	4	BJ458684	BJ458684
C	14	41	93.2	487	4	BJ448746	BJ448746
C	15	41	93.2	482	4	BJ447959	BJ447959
C	16	41	93.2	490	4	BJ451181	BJ451181
C	17	41	93.2	502	6	CD057466	CD057466
C	18	41	93.2	504	1	AJ434780	AJ434780
C	19	41	93.2	536	7	CK570032	CK570032
C	20	41	93.2	554	2	BE704839	BE704839
C	21	41	93.2	573	6	CB877974	CB877974
C	22	41	93.2	590	2	BE492776	BE492776
C	23	41	93.2	600	2	BE705635	BE705635
C	24	41	93.2	616	1	AV909381	AV909381
C	25	41	93.2	619	4	BJ460127	BJ460127
C	26	41	93.2	628	6	CB878395	CB878395
C	27	41	93.2	640	1	AV911273	AV911273
C	28	41	93.2	643	4	BJ476744	BJ476744
C	29	41	93.2	648	6	CB877648	CB877648
C	30	41	93.2	652	4	BJ452579	BJ452579
C	31	41	93.2	676	4	BJ453684	BJ453684
C	32	41	93.2	683	4	BJ458193	BJ458193
C	33	41	93.2	706	4	BJ461226	BJ461226
C	34	41	93.2	711	2	BE214392	BE214392
C	35	41	93.2	725	2	BF623796	BF623796
C	36	41	93.2	731	4	BG300006	BG300006
C	37	41	93.2	739	2	BF622309	BF622309
C	38	41	93.2	741	2	BF621719	BF621719
C	39	41	93.2	749	2	BF623605	BF623605
C	40	41	93.2	796	2	BF624602	BF624602
C	41	41	93.2	1101	9	CNS04VYJ	CNS04VYJ
C	42	41	93.2	1175	2	BE421167	BE421167
C	43	39	88.6	439	8	AZ591722	AZ591722
C	44	39	88.6	485	4	BJ348039	BJ348039
C	45	39	88.6	510	1	AL638926	AL638926

#### ALIGNMENTS

RESULT 1  
LOCUS BJ457822 150 bp mRNA linear EST 23-MAY-2002  
DEFINITION BJ457822 K. Sato unpublished cDNA library, cv. Akashinriki  
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone  
Baak29g22.3', mRNA sequence.  
ACCESSION BJ457822.1 GI:21136365  
VERSION BJ457822  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
1 (bases 1 to 150)

REFERENCE  
AUTHORS Sato, K., Saitoh, D. and Takeda, K.  
TITLE Barley EST sequencing project in NIG and Okayama Univ  
JOURNAL Unpublished (2002)  
COMMENT Contact: Tadasu Shin-1  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@gene.nig.ac.jp.  
Location/Qualifiers

#### FEATURES

source  
1..150  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultiivar="Akashinriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="Baak29g22"

/cissue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_1ib="K. Sato unpublished cDNA library, cv.  
Akashiniriki vegetative stage leaves"

## ORIGIN

## Alignment Scores:

Pred. No.:	154	Length:	150
Score:	41.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:	4	Gaps:	0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x BJ457822 (1-150)

QY 1 MechH1eua1aleuProAlaSerA 9

DB 93 CTTGATTGGCGCTGCCGCGCTGCGCT 119

## RESULT 2

BJ474645

LOCUS BJ474645 K. Sato unpublished cDNA library, cv. Haruna NiJo adult,  
heading stage top three leaves Hordeum vulgare subsp. vulgare cDNA

DEFINITION clone baal20g09 3', mRNA sequence.

ACCESSION BJ474645 GI:21153146

VERSION BJ474645

KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare

ORGANISM Hordeum vulgare subsp. vulgare

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.

1 (bases 1 to 181)

AUTHORS Sato, K., Saitoh, D. and Takeda, K.

TITLE Barley EST sequencing project in NIG and Okayama Univ

JOURNAL Unpublished (2002)

COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers

FEATURES

source  
1..181  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Haruna NiJo"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baal20g09"  
/cissue\_type="top three leaves"  
/dev\_stage="adult, heading stage"  
/clone\_1ib="K. Sato unpublished cDNA library, cv. Haruna  
NiJo adult, heading stage top three leaves"

ORIGIN

## Alignment Scores:

Pred. No.:	189	Length:	181
Score:	41.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:	4	Gaps:	0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x BJ474645 (1-181)

QY 1 MechH1eua1aleuProAlaSerA 9

DB 89 CTTGATTGGCGCTGCCGCGCTGCGCT 115

## RESULT 3

BJ450407/c

LOCUS BJ450407 K. Sato unpublished cDNA library, cv. Akashiniriki  
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone

DEFINITION baak29g22 5', mRNA sequence.

ACCESSION BJ450407

VERSION BJ450407.1 GI:21129011

KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare

ORGANISM Hordeum vulgare subsp. vulgare

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.

1 (bases 1 to 202)

AUTHORS Sato, K., Saitoh, D. and Takeda, K.

TITLE Barley EST sequencing project in NIG and Okayama Univ

JOURNAL Unpublished (2002)

COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers

FEATURES

source  
1..202  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Akashiniriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baak29g22"  
/cissue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_1ib="K. Sato unpublished cDNA library, cv.  
Akashiniriki vegetative stage leaves"

ORIGIN

Alignment Scores:

Pred. No.:	213	Length:	202
Score:	41.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:	4	Gaps:	0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x BJ450407 (1-202)

QY 1 MechH1eua1aleuProAlaSerA 9

DB 49 CTTGATTGGCGCTGCCGCGCTGCGCT 23

## RESULT 4

CD662805

LOCUS UCRHV18\_03ag11 b1 Drought-stressed Dicktoo barley epidermis cDNA  
library Hordeum vulgare subsp. vulgare cDNA clone UCRHV18\_03ag11,  
mRNA sequence.

DEFINITION CD662805 234 bp mRNA linear EST 23-JUN-2003

ACCESSION CD662805

VERSION CD662805.1 GI:32153419

KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare

ORGANISM Hordeum vulgare subsp. vulgare

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.

1 (bases 1 to 234)

AUTHORS Wahid, A., Close, T. J., Fenton, R. D., Wanamaker, S., Collura, K.,  
Feuerbacher, O., Kim, H. R., Kudrna, D., Wing, R. and Yu, Y.

TITLE Drought-stressed barley leaf epidermis cDNA sequences

JOURNAL Unpublished (2003)

COMMENT Contact: Timothy J. Close

Department of Botany & Plant Sciences  
University of California  
Riverside, CA 92521-0124, USA  
Tel: 909-787-3318  
Fax: 909-787-4437  
Email: timothy.close@ucr.edu  
Seq primer: T7.

# FEATURES

source

Location/Qualifiers  
1..234  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Dicktoo"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="UCRHV18\_03ag11"  
/tissue\_type="lower leaf epidermis"  
/dev\_stage="1-2 week seedlings"  
/lab\_host="E. coli TUC121"  
/clone\_idb="Drought-stressed Dicktoo barley epidermis cDNA library"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid:  
Site\_1: EcoRI; Site\_2: XhoI; Seeds of barley (Hordeum vulgare L. cv. Dicktoo) were germinated in dispanes containing UCR-mix soil. Seedlings were kept in a growth chamber at 20C (day/night) and allowed to grow at 60-70% soil moisture content. After 3-4 days, the water was withheld in order to apply drought until the soil moisture content was reduced to 10-12%, which took another 3-4 days. At the time of extraction of epidermis, leaf water and osmotic potentials and extracted lower epidermis osmotic potentials had dropped to 15.9%, 20.2% and 24.7% of controls, respectively, as determined using a vapor pressure osmometer (Model 5100C, Wescor Inc., Logan, UT). Epidermal strips were quickly peeled off of seedlings and immediately submerged in liquid nitrogen. About 15-20 g of epidermal tissue was collected and used to extract total RNA. Total RNA was extracted using Concert Plant RNA Reagent (Invitrogen 12322-012). Poly(A) RNA was purified using PolyAtrack mRNA isolation system IV (Promega). A cDNA library was made using a Uni-ZAP cDNA synthesis kit (Stratagene). A total of 1 million primary lambda cDNA clones were mass-excised in vivo to give a population of phagescript SK(-) phagemids. All steps to this point were performed in the TJ Close lab at the University of California, Riverside by A. Mahid with some assistance from R.D. Fenton. Phagemids were plated on the TUC121 host strain, plasmid DNA purified, cDNA clones archived, and DNA sequences determined using the T7 primer (mainly 3' end reads) using an ABI3730 at the Arizona Genomics Institute, University of Arizona (Collura, Feuerbacher, Kim, Kudrna, Ming, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wanmaker) using the Harvest pipeline (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to Genbank."

## Alignment Scores:

Pred. No.:	251	Length:	234
Score:	41.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:	6	Gaps:	0

## ORIGIN

US-10-617-443b-2\_COPY\_1\_9 (1-9) x CD662805 (1-234)

QY 1 MechisLeuAlaLeuProAlaSerAla 9  
DB 195 CTCATTGGCGCTGCGCGCTGCGCT 221

## RESULT 5

AV912195

LOCUS

DEFINITION

AV912195 308 bp mRNA linear EST 18-JAN-2002  
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone baak1f21 3', mRNA sequence.

ACCESSION

AV912195

VERSION

AV912195.1

KEYWORDS:

EST

SOURCE

Hordeum vulgare subsp. vulgare  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum

REFERENCE

Sato, K., Salsino, D. and Takeda, K.  
Barley EST sequencing project in NIG and Okayama Univ  
Unpublished (2002)  
Contact: Tadasu Shin-i  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tahin@genes.nig.ac.jp.

## FEATURES

source

Location/Qualifiers  
1..308  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Akashinriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baak1f21"  
/tissue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_idb="K. Sato unpublished cDNA library, cv. Akashinriki vegetative stage leaves"

## ORIGIN

### Alignment Scores:

Pred. No.:	339	Length:	308
Score:	41.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:	1	Gaps:	0

US-10-617-443b-2\_COPY\_1\_9 (1-9) x AV912195 (1-308)

QY 1 MechisLeuAlaLeuProAlaSerAla 9  
DB 133 CTCATTGGCGCTGCGCGCTGCGCT 159

## RESULT 6

CD662622

LOCUS

DEFINITION

CD662622 314 bp mRNA linear EST 23-JUN-2003  
UCRHV18\_02cc08.b1 Drought-stressed Dicktoo barley epidermis cDNA library Hordeum vulgare subsp. vulgare cDNA clone UCRHV18\_02cc08, mRNA sequence.

ACCESSION

CD662622

VERSION

CD662622.1

KEYWORDS

EST

SOURCE

Hordeum vulgare subsp. vulgare  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum

REFERENCE

Wahid, A., Close, T.J., Fenton, R.D., Wanmaker, S., Collura, K.,  
Feuerbacher, O., Kim, H.R., Kudrna, D., Ming, R. and Yu, Y.  
Drought-stressed barley leaf epidermis cDNA sequences  
Unpublished (2003)  
Contact: Timothy J. Close

Department of Botany & Plant Sciences  
University of California  
Riverside, CA 92521-0124, USA  
Tel: 909-787-3318  
Fax: 909-787-4437  
Email: timothy.cloose@ucr.edu

Seq primer: T7.  
Location/Qualifiers

1.314  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="rRNA"  
/cultiivar="Dicktoo"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="UCRHY18.02cc08"  
/tissue\_type="lower leaf epidermis"  
/dev\_stage="1-2 week seedlings"  
/lab\_host="E. coli TUC121"  
/clone\_lib="Drought-stressed Dicktoo barley epidermis cDNA library"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site\_1: EcoRI; Site\_2: XhoI; Seeds of barley (Hordeum  
vulgare L. cv. Dicktoo) were germinated in diaphans  
containing UCR-mix soil. Seedlings were kept in a growth  
chamber at 20C (day/night) and allowed to grow at 60-70%  
soil moisture content. After 3-4 days, the water was  
withheld in order to apply drought until the soil moisture  
content was reduced to 10-12%, which took another 3-4  
days. At the time of extraction of epidermis, leaf water  
and osmotic potentials and extracted lower epidermis  
osmotic potentials had dropped to 15.9%, 20.2% and 24.7%  
of controls, respectively. As determined using a vapor  
pressure osmometer (Model 5100C, Wescor, Inc., Logan, UT).  
Epidermal strips were quickly peeled off of seedlings and  
immediately submerged in liquid nitrogen. About 15-20 g of  
epidermal tissue was collected and used to extract total  
RNA. Total RNA was extracted using Concert Plant RNA  
Reagent (Invitrogen 12322-012). Poly(A) RNA was purified  
using PolyAtract mRNA Isolation System IV (Promega). A  
cDNA library was made using a Uni-ZAP cDNA synthesis kit  
(Stratagene). A total of 1 million primary lambda cDNA  
clones were mass-excised in vivo to give a population of  
phagescript SK(-) phagemids. All steps to this point were  
performed in the Tj Close lab at the University of  
California, Riverside by A. Mahd with some assistance  
from R.D. Fenton. Phagemids were plated on the TUC121 host  
strain, plasmid DNA purified, cDNA clones archived, and  
DNA sequences determined using the T7 primer (mainly 3'  
end reads) using an ABI3730 at the Arizona Genomics  
Institute, University of Arizona (Collura, Feuerbacher,  
Kim, Kudrna, Wing, Yu). Chromatogram files were  
transmitted to UC Riverside (by Yu), then processed at UC  
Riverside (by Manamaker) using the HarVEST pipeline  
(http://harvest.ucr.edu) to remove vector and cloning  
oligo sequences and various contaminants, and to trim to a  
high quality region. Sequences that remained a pired 17  
region of at least 100 bases were deposited to GenBank."

#### ORIGIN

Alignment Scores:  
Pred. No.: 346 Length: 314  
Score: 41.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 93.18% Indels: 0  
DB: 6 Gaps: 0

US-10-617-443b-2\_COPY\_1\_9 (1-9) x CD662622 (1-314)

QY 1 MethiLeuAlaLeuProAlaSerAla 9  
Db 177 CTTCAATTGGCGCTGCGCGCTGCGCT 203

RESULT 7  
AV0909564/c 329 bp mRNA linear EST 18-JAN-2002  
LOCUS  
DEFINITION  
AV0909564 K. Sato unpublished cDNA library, cv. Akashinriki  
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone  
baak13601 5', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Hordeum vulgare subsp. vulgare  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poaceae; Triticeae; Hordeum.  
1 (bases 1 to 329)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Sato, K., Saitoh, D. and Takeda, K.  
Barley EST sequencing project in NIG and Okayama Univ  
Unpublished (2002)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tsuhigene@nig.ac.jp.

FEATURES  
source  
1..329  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="rRNA"  
/cultiivar="Akashinriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baak13601"  
/tissue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_lib="K. Sato unpublished cDNA library, cv.  
Akashinriki vegetative stage leaves"

#### ORIGIN

Alignment Scores:  
Pred. No.: 365 Length: 329  
Score: 41.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 93.18% Indels: 0  
DB: 1 Gaps: 0

US-10-617-443b-2\_COPY\_1\_9 (1-9) x AV0909564 (1-329)

QY 1 MethiLeuAlaLeuProAlaSerAla 9  
Db 185 CTTCAATTGGCGCTGCGCGCTGCGCT 159

RESULT 8  
AV0911456 332 bp mRNA linear EST 18-JAN-2002  
LOCUS  
DEFINITION  
AV0911456 K. Sato unpublished cDNA library, cv. Akashinriki  
vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone  
baak13601 3', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Hordeum vulgare subsp. vulgare  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poaceae; Triticeae; Hordeum.  
1 (bases 1 to 332)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Sato, K., Saitoh, D. and Takeda, K.  
Barley EST sequencing project in NIG and Okayama Univ  
Unpublished (2002)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information

National Institute of Genetics  
111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.

FEATURES  
source  
1. .332  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Akashinriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="bak3908"  
/issue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_lib="K. Sato unpublished cDNA library, cv. Akashinriki vegetative stage leaves"

ORIGIN

Alignment Scores:  
Pred. No.: 368 Length: 332  
Score: 41.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 93.18% Indels: 0  
DB: 1 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x AV911456 (1-332)

Qy 1 MechisLeuAlaleuProAlaSerAla 9  
Db 131 CTCATTGGCGCTGCCGCGCTCGCT 157

RESULT 9  
LOCUS AV912337 419 bp mRNA linear EST 18-JAN-2002  
DEFINITION AV912337 K. Sato unpublished cDNA library, cv. Akashinriki vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone bak3908 3', mRNA sequence.

ACCESSION AV912337 GI:18208114  
VERSION AV912337.1  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE 1 (bases 1 to 419)  
Sato, K., Saitoh, D. and Takeda, K.  
Barley EST sequencing project in NIG and Okayama Univ  
Unpublished (2002)  
Contact: Tadao Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.

FEATURES  
source  
1. .419  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Akashinriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="bak3908"  
/issue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_lib="K. Sato unpublished cDNA library, cv. Akashinriki vegetative stage leaves"

ORIGIN

Alignment Scores:

Pred. No.: 476 Length: 419  
Score: 41.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 93.18% Indels: 0  
DB: 1 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x AV912337 (1-419)

Qy 1 MechisLeuAlaleuProAlaSerAla 9  
Db 132 CTCATTGGCGCTGCCGCGCTCGCT 158

RESULT 10  
LOCUS AV910159 427 bp mRNA linear EST 18-JAN-2002  
DEFINITION AV910159 K. Sato unpublished cDNA library, cv. Akashinriki vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone bak3908 5', mRNA sequence.

ACCESSION AV910159 GI:18205901  
VERSION AV910159  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE 1 (bases 1 to 427)  
Sato, K., Saitoh, D. and Takeda, K.  
Barley EST sequencing project in NIG and Okayama Univ  
Unpublished (2002)  
Contact: Tadao Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.

FEATURES  
source  
1. .427  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Akashinriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="bak3908"  
/issue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_lib="K. Sato unpublished cDNA library, cv. Akashinriki vegetative stage leaves"

ORIGIN

Alignment Scores:  
Pred. No.: 486 Length: 427  
Score: 41.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 93.18% Indels: 0  
DB: 1 Gaps: 0

US-10-617-443B-2\_COPY\_1\_9 (1-9) x AV910159 (1-427)

Qy 1 MechisLeuAlaleuProAlaSerAla 9  
Db 289 CTCATTGGCGCTGCCGCGCTCGCT 263

RESULT 11  
LOCUS BJ461882 440 bp mRNA linear EST 23-MAY-2002  
DEFINITION BJ461882 K. Sato unpublished cDNA library, cv. Akashinriki vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone bak21b2 3', mRNA sequence.

VERSION BJ461882.1 GI:21140391  
 KEYWORDS EST  
 SOURCE Hordeum vulgare subsp. vulgare  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticeae; Hordeum.  
 1 (bases 1 to 440)  
 REFERENCE Sato, K., Saitoh, D. and Takeda, K.  
 AUTHORS Barley EST sequencing project in NIG and Okayama Univ  
 TITLE Unpublished (2002)  
 JOURNAL Contact: Tadasu Shin-i  
 COMMENT Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.  
 Location/Qualifiers  
 1..440  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Akashinriki"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="baak21b12"  
 /tissue\_type="leaves"  
 /dev\_stage="vegetative stage"  
 /clone\_lib="K. Sato unpublished cDNA library, cv.  
 Akashinriki vegetative stage leaves"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 502 Length: 440  
 Score: 41.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 93.18% Indels: 0  
 Gaps: 0  
 DB: 4

US-10-617-443B-2\_COPY\_1\_9 (1-9) x BJ461882 (1-440)  
 QY 1 Meth1sleuAlaleuProAlaSerAla 9  
 :::::::::::::::::::::::::::::::  
 DB 102 CTTCAATTGGCGCTGCCGCTCGCT 128

RESULT 12  
 BJ455746 446 bp mRNA linear EST 23-MAY-2002  
 LOCUS BJ455746 K. Sato unpublished cDNA library, cv. Akashinriki  
 DEFINITION vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone  
 baak18a21 3', mRNA sequence.  
 ACCESSION BJ455746  
 VERSION BJ455746.1 GI:21134312  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. vulgare  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticeae; Hordeum.  
 1 (bases 1 to 446)  
 REFERENCE Sato, K., Saitoh, D. and Takeda, K.  
 AUTHORS Barley EST sequencing project in NIG and Okayama Univ  
 TITLE Unpublished (2002)  
 JOURNAL Contact: Tadasu Shin-i  
 COMMENT Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.  
 Location/Qualifiers  
 1..446

ORIGIN  
 Alignment Scores:  
 Pred. No.: 509 Length: 446  
 Score: 41.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 93.18% Indels: 0  
 Gaps: 0  
 DB: 4

US-10-617-443B-2\_COPY\_1\_9 (1-9) x BJ455746 (1-446)  
 QY 1 Meth1sleuAlaleuProAlaSerAla 9  
 :::::::::::::::::::::::::::::::  
 DB 104 CTTCAATTGGCGCTGCCGCTCGCT 130

RESULT 13  
 BJ458684 469 bp mRNA linear EST 23-MAY-2002  
 LOCUS BJ458684 K. Sato unpublished cDNA library, cv. Akashinriki  
 DEFINITION vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone  
 baak33a14 3', mRNA sequence.  
 ACCESSION BJ458684  
 VERSION BJ458684.1 GI:21137220  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. vulgare  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticeae; Hordeum.  
 1 (bases 1 to 469)  
 REFERENCE Sato, K., Saitoh, D. and Takeda, K.  
 AUTHORS Barley EST sequencing project in NIG and Okayama Univ  
 TITLE Unpublished (2002)  
 JOURNAL Contact: Tadasu Shin-i  
 COMMENT Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.  
 Location/Qualifiers  
 1..469  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Akashinriki"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="baak33a14"  
 /tissue\_type="leaves"  
 /dev\_stage="vegetative stage"  
 /clone\_lib="K. Sato unpublished cDNA library, cv.  
 Akashinriki vegetative stage leaves"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 538 Length: 469  
 Score: 41.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 93.18% Indels: 0  
 Gaps: 0  
 DB: 4



```

US-10-617-443B-2_COPY_1_9 (1-9) x BJ458684 (1-469)
Cy      1 MethistenualeuProAlaSerAla 9
Db      126 CTTCAATTGGCGCTGCCGCTCGCT 152

RESULT 14
BJ448746      482 bp mRNA linear EST 23-MAY-2002
LOCUS      BJ448746 K. Sato unpublished cDNA library, cv. Akashinriki
DEFINITION      vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
                baak2jb12 5', mRNA sequence.
ACCESSION      BJ448746
VERSION      BJ448746
KEYWORDS      EST.
SOURCE      Hordeum vulgare subsp. vulgare
ORGANISM      Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 482)
Sato, K., Saitoh, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..482
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Akashinriki"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="baak2jb12"
/tissue_type="leaves"
/dev_stage="vegetative stage"
/clone_lib="K. Sato unpublished cDNA library, cv.
Akashinriki vegetative stage leaves"

ORIGIN
Alignment Scores:
Pred. No.: 555 Length: 482
Score: 41.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 93.18% Indels: 0
DB: 4 Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x BJ448746 (1-482)
Cy      1 MethistenualeuProAlaSerAla 9
Db      338 CTTCAATTGGCGCTGCCGCTCGCT 312

RESULT 15
BJ447999      487 bp mRNA linear EST 23-MAY-2002
LOCUS      BJ447999 K. Sato unpublished cDNA library, cv. Akashinriki
DEFINITION      vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
                baak18a21 5', mRNA sequence.
ACCESSION      BJ447999
VERSION      BJ447999
KEYWORDS      EST.
SOURCE      Hordeum vulgare subsp. vulgare
ORGANISM      Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.

```

```

REFERENCE      1 (bases 1 to 487)
AUTHORS      Sato, K., Saitoh, D. and Takeda, K.
TITLE      Barley EST sequencing project in NIG and Okayama Univ
JOURNAL      Unpublished (2002)
COMMENT      Contact: Tadasu Shin-i
                Center For Genetic Resource Information
                National Institute of Genetics
                1111 Yata, Mishima, Shizuoka 411-8540, Japan
                Tel: 81-559-81-6856
                Fax: 81-559-81-6855
                Email: tshini@genes.nig.ac.jp.
                Location/Qualifiers
1..487
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Akashinriki"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="baak18a21"
/tissue_type="leaves"
/dev_stage="vegetative stage"
/clone_lib="K. Sato unpublished cDNA library, cv.
Akashinriki vegetative stage leaves"

ORIGIN
Alignment Scores:
Pred. No.: 561 Length: 487
Score: 41.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 93.18% Indels: 0
DB: 4 Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x BJ447999 (1-487)
Cy      1 MethistenualeuProAlaSerAla 9
Db      343 CTTCAATTGGCGCTGCCGCTCGCT 317

Search completed: June 30, 2005, 01:27:45
Job time : 161.519 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_p2n model

Run on: June 29, 2005, 19:59:10 ; Search time 7513.54 Seconds

(without alignments):  
2153.984 Million cell updates/sec

Title: US-10-617-443B-2

Perfect score: 1679

Sequence: 1 MHIALPASAGLHQLSPRYK.....LEVRGRNDLLFLSLAEVVM 334

Scoring table:  
BIOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xjh  
-Q=/cgnr\_1/USPTO.spool/US10617443/runat\_29062005\_135931\_26061/app\_query.fasta\_1.718  
-DB=Genemb1 -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USFR=US10617443 @CGN\_1\_1\_3590 @runat\_29062005\_135931\_26061 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Genemb1.\*  
1: gb Da.\*  
2: gb Hrg.\*  
3: gb In.\*  
4: gb Om.\*  
5: gb Ov.\*  
6: gb Pat.\*  
7: gb Ph.\*  
8: gb Pl.\*  
9: gb Pr.\*  
10: gb Ro.\*  
11: gb Sts.\*  
12: gb Sy.\*  
13: gb Un.\*  
14: gb Vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1635	97.4	1797 6	AX338509 Sequence
2	1635	97.4	2040 6	BD260123 Matrix-re
3	1635	97.4	2543 9	AY280665 Homo sapi
4	1635	97.4	2554 9	AY040094 Homo sapi

5	1635	97.4	2585	9	BC034390	BC034390 Homo sapi
6	1635	97.4	2586	9	BC035717	BC035717 Homo sapi
7	1536	91.5	2446	10	AY156509	AY156509 Mus muscu
8	1524	90.8	1338	6	CO879976	CO879976 Sequence
9	1463.5	87.2	2525	10	AY037300	AY037300 Mus muscu
10	1450.5	86.4	2007	5	CR353995	CR353995 Gallus ga
11	1450.5	86.4	2176	5	BX929862	BX929862 Gallus ga
12	1238.5	73.8	2080	5	BC078402	BC078402 Dario rer
13	1117	66.5	1953	9	AY280666	AY280666 Homo sapi
14	1102.5	65.7	819	6	CO716317	CO716317 Sequence
15	1057	63.0	1897	10	AY280664	AY280664 Mus muscu
16	1042.5	62.1	1113	4	AF097707	AF097707 Bos tauru
17	1034.5	61.6	1894	4	AX747270	AX747270 Sequence
18	1034.5	61.6	1894	9	AK031944	AK031944 Homo sapi
19	1034.5	61.6	1900	9	AK032476	AK032476 Homo sapi
20	1034.5	61.6	2029	9	HSSERIRPT	Y07971 H.sapiens m
21	1034.5	61.6	2036	6	CO728266	CO728266 Sequence
22	1034.5	61.6	2036	6	D87258	D87258 Homo sapien
23	1034.5	61.3	2075	6	E12965	E12965 cDNA encodi
24	1029.5	61.3	1169	9	AF097709	AF097709 Homo sapi
25	1029.5	61.3	2043	9	BC031082	BC031082 Homo sapi
26	1027.5	61.2	2012	10	AF179370	AF179370 Rattus no
27	1027.5	61.2	2036	6	AK035627	AK035627 Sequence
28	1027.5	61.2	2065	10	BC081767	BC081767 Rattus no
29	1026.5	61.1	1777	5	BX934040	BX934040 Gallus ga
30	1026.5	61.1	2005	10	AF179369	AF179369 Mus muscu
31	1026.5	61.1	2007	10	AK128916	AK128916 Mus muscu
32	1023.5	61.0	2023	10	BC013516	BC013516 Mus muscu
33	1006.5	59.9	1705	6	BD194550	BD194550 Human nuc
34	1006.5	59.9	1705	6	AX015404	AX015404 Sequence
35	1000.5	59.6	2010	10	AF172994	AF172994 Mus muscu
36	989.5	58.9	2055	5	BC074069	BC074069 Dario rer
37	981.5	58.5	1534	6	AX020200	AX020200 Sequence
38	925	55.1	1362	10	BC022646	BC022646 Mus muscu
39	893.5	53.2	915	5	BX950476	BX950476 Gallus ga
40	890	53.0	1544	6	CO783299	CO783299 Sequence
41	890	53.0	1544	6	BD127576	BD127576 Primer fo
42	890	53.0	1544	9	AK075205	AK075205 Homo sapi
43	890	53.0	1553	6	BD063243	BD063243 Secreted
44	888	52.9	1521	9	BC057765	BC057765 Homo sapi
45	883	52.6	1625	10	AF164513	AF164513 Mus muscu

## ALIGNMENTS

RESULT 1	AX338509	Sequence 28 from Patent WO0183775.	1797 bp	DNA	linear	PAT 09-JAN-2002
LOCUS	AX338509					
DEFINITION	AX338509					
ACCESSION	AX338509.1	GI:18128931				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
	Homo sapiens (human)					
	Homo sapiens					
REFERENCE						
AUTHORS	Deleageane,A.M., Lal,P., Hafalia,A., Patterson,C., Walla,N.K., Kearney,L., Tribouley,C.M., Khan,F.A., Yao,M.G., Baughn,M.R., Azimzai,Y., Elliott,V.S., Nguyen,D.B., Gandhi,A.R., Yang,J., Hernandez,R., Policky,J.L., Lu,D.A., Reddy,R., and Tang,Y.T.					
TITLE	Proteases					
JOURNAL	Patent: WO 0183775-A 28 08-NOV-2001;					
FEATURES	Incyle Genomics, Inc. (US)					
source	1. 1797					
	/organism="Homo sapiens"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:9606"					
	/note="Incyle ID No: 7474343CB1"					
ORIGIN						
Alignment Scores:						

Pred. No.:	1.06e-130	Length:	1797
Score:	1635.00	Matches:	325
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	97.38%	Gaps:	0
DB:	6		
US-10-617-443b-2 (1-334) x AX338509 (1-1797)			
QY	10	GlyLeuH1eGlnLeuSerSerProAgtTylLysPheAenPheH1eAlaAapValValGlu	29
DB	565	GGTCCTCCACGAGTCAGACACCCGCGCTACAGTTCATCTGTCGACGGTGGAG	624
QY	30	LysIleAlaProAlaValValH1eGluPheLeuAgtH1sProLeuPheG1YArg	49
DB	625	AAAGTCGACACGCGGTGTCACATAGACTCTTCTTGAGACCCGCTGTTGGCCG	684
QY	50	AenValProLeuSerSerGlySerGlyPheIleMetSerGluAglYLeuIleIleThr	69
DB	685	AAAGTCGCGCTGTCAGCGGTTCTGCTCATGTCAGAGCCGCGCTGATCATCACC	744
QY	70	AenAlaH1sValValSerSerAenSerAlaAlaProGlyArgGlnGluLeuValGln	89
DB	745	AAAGCCACGAGTGTCTCAGACCAAGTGTGCCCCGCGCAGCAGCATCAAGTGCAG	804
QY	90	LeuGlnAenGlyAenSerTyrGluAlaThrIleLysAenIleAenAplyLysSerAply	109
DB	805	CTTCAGAGTGGGAGCTCTTAGAGGCCACATCAAGACATCAAGAGAGTGGACATT	864
QY	110	AlaThrIleLysIleH1sProLysLysLeuProValLeuLeuLeuGlnH1sSerAla	129
DB	865	GCCACCATCAAGATCCATCCCAAGAAAGCTCCCTGTGTGTGTGGTGGGTCACTGGCC	924
QY	130	AapLeuAgtProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAenThr	149
DB	925	GACCTGGCGCTGGAGGTTGTGGGCGCATGGCAGTCCCTTCCCTCAAGAACACA	984
QY	150	ValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuGlyLysArgA	169
DB	985	GTCACAGACGGGACCTGTCACACTGCCACCGGAGGCGAGGAGCTGGGCTCCGGGAC	1044
QY	170	SerAapMetAapTyrIleGlnThrAapAlaIleIleAenTyrGlyAenSerGlyGlyPro	189
DB	1045	TCCGACATGAGTACATCCAGACGAGTGCATCATCACTCAAGGAACTCCGGGGACCA	1104
QY	190	LeuValAenLeuAapGlyGluValIleGlyIleAenThrLeuLysValThrAlaGlyIle	209
DB	1105	CTGGTGAACCTGATGGCGAGTCATTTGGCATCAACGCTCAAGGTCAAGCTGGCATC	1164
QY	210	SerPheAlaIleProSerAapArgIleThrArgPheLeuThrGluPheGlnAapLysGln	229
DB	1165	TCTTTGGCATCCCTTCAGACCGCATCACCGGTTCTCTCAAGATTTCCAGACAGCAG	1224
QY	230	IleLysAapTyrLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu	249
DB	1225	ATCAAAAGACTGGAAGAAGCGCTTCATCGGCATCGAGTGGAGCGATCACCAAGCCTG	1284
QY	250	ValAapGlyLeuLysLysAenProAapPheProGluValSerSerGlyIleTyrVal	269
DB	1285	GTCGATGAGCTGAAGCGCAGCAACCGGACTTCCAGAGGTGAGCTGAATTTATGAG	1344
QY	270	GlnGlnValAlaProAenSerProSerGlnArgGlyGlyIleGlnAapGlyAapIleIle	289
DB	1345	CAGAGAGTGGCGCGAATTCATCTTCCAGAGAGGCGCATCCAGATGGTGCATCATC	1404
QY	290	ValLysValAenGlyArgProLeuValAapSerSerGlyLeuGlnGluValLeuThr	309
DB	1405	GTCAGAGTCAAGGGGCTCTCTAGTGAAGTGAAGTGAAGGCGCGCTGATCACC	1464
QY	310	GluSerProLeuLeuLeuGluValArgArgGlyAenAapAapLeuLeuPheSerIleAla	329
DB	1465	GAGTCTCTCTCTACTGAGGTGGCGGGGAGAACAGACGACTCTTCCAGCATCGCA	1524

QY	330	ProGluValValMet	334
DB	1525	CCTGAGGTGGTCAATG	1539
RESULT 2			
LOCUS	BD260123		
DEFINITION	Matrix-remodeling genes.	2040 bp	DNA
ACCESSION	BD260123		
VERSION	BD260123.1 GI:33069893		
KEYWORDS	JP 2002527054-A/14.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2040)		
TITLE	Walker,M.G., Volkmutch,W. and Klingler,T.M.		
JOURNAL	Matrix-remodeling genes Patent: JP 2002527054-A 14 27-AUG-2002;		
COMMENT	INCYTE PHARMACEUTICALS INC OS Homo sapiens (human)		
PN	JP 2002527054-A/14		
PD	27-AUG-2002		
PF	06-OCT-1999 JP 2000575891		
PR	09-OCT-1998 US 09/169289		
PI	MICHAEL G WALKER, WAYNE VOLKUTCH, TOD M KLINGLER PC		
PC	CI2N15/09, A61K38/00, A61K38/28, A61K38/46, A61K48/00, A61P1/04, PC		
PC	A61P3/10, A61P9/10, A61P17/02, A61P19/02, A61P35/00, C07K14/47, C07K16/18, PC		
PC	CI2N1/15,		
PC	CI2N1/19, CI2N1/21, CI2N5/10, CI2Q1/68, G01N33/15, G01N33/50, G01N33/ PC		
53,	G01N33/53//A61K31/711, A61K35/76, CI2N15/00, CI2N5/00, A61K37/02,		
PC	A61K37/26,		
PC	A61K37/54		
CC	Incyte ID NO.: 2814981CB1		
FT	Key Location/Qualifiers		
FT	source 1..2040		
FT	Location/Qualifiers		
FEATURES	1..2040 /organism='Homo sapiens (human)'		
source	/organism='Homo sapiens'		
	/mol_type='Genomic DNA'		
	/db_xref='taxon:9606'		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.25e-130	Length:	2040
Score:	1635.00	Matches:	325
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.38%	Indels:	0
DB:	6	Gaps:	0
US-10-617-443b-2 (1-334) x BD260123 (1-2040)			
QY	10	GlyLeuH1eGlnLeuSerSerProAgtTylLysPheAenPheH1eAlaAapValValGlu	29
DB	72	GGTCCTCCACGAGTCAGACACCCGCGCTACAGTTCATCTTGTGAGTGGAG	131
QY	30	LysIleAlaProAlaValValH1eGluPheLeuAgtH1sProLeuPheG1YArg	49
DB	132	AAAGTCGACACGCGGTGTCACATAGACTCTTCTTGAGACACCCGCTGTTGGCCG	191
QY	50	AenValProLeuSerSerGlySerGlyPheIleMetSerGluAglYLeuIleIleThr	69
DB	192	AAAGTCCCGCTGTCCAGCGGTTCTGCTTCATGTCAGAGGCGCGGCTGATCATCACC	251
QY	70	AenAlaH1sValValSerSerAenSerAlaAlaProGlyArgGlnGluLeuValGln	89
DB	252	AAAGCCACGAGTGTCTCAGACCAAGTGTGCTCCCGGGCAGCAGCATCAAGTGCAG	311

Qy 90 LeuGlnbngiYAspSerTYrGluAlaThrIleYAspIleAspLYsSerAspIle 109  
 Db 312 CTACGAGATGGGAGCTCTTAGAGCCCAATCAAGCATCGACAAAGATCGGACATT 371  
 Qy 110 AlaThrIleYsIleHisProLYsLYsLeuProValLeuLeuGlyHisSerAla 129  
 Db 372 GCCACCATCAAGATCCATCCCAAGAAAAGCTCCCTGTTGTTGCTGGTCACTCGGCC 431  
 Qy 130 AspLeuArgProGlyYgluPheValValAlaIleGlySerProPheAlaLeuGlnAspThr 149  
 Db 432 GACCTGGGCTGGGAGATTGTGTGGCATCGGAGTCCCTTCGCTTACAGAACCA 491  
 Qy 150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAsp 169  
 Db 492 GTGACAAACGGGACATGTCAGCATGCGCAGGGAGGCAAGGAGCTGGCTCCGGGAC 551  
 Qy 170 SerAspMetAspTYrIleGlnThrAspAlaIleIleAsnTYrGlyAsnSerGlyGlyPro 189  
 Db 552 TCCGACATGACATCATCCAGACGAGTCCATCATCACTACGGGAACTCCGGGGAGACA 611  
 Qy 190 LeuValAsnLeuAspGlyYgluValIleGlyIleAsnThrLeuYsValThrAlaGlyIle 209  
 Db 612 CTGGTGAACCTGGATGGGAGGATGATGATGATCAACAGCTCAAGGTCACGGCTGGCATC 671  
 Qy 210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLYsGln 229  
 Db 672 TCCCTTCCATCCCTCGACCGCATCAAGGTTCTCTCAAGAGTTCCAAAGCAAGCAG 731  
 Qy 230 IleYsAspTYrLYsLYsArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249  
 Db 732 ATCAAAAGCTGGAAGAGCGCTTCATCGGCATACGATGCGGACGATCAACAGCCTG 791  
 Qy 250 ValAspGluLeuYsAlaSerAspProAspPheProGluValSerSerGlyIleTYrVal 269  
 Db 792 GTGATAGAGCGAAGGCGACAGACCCGAGCTTCCAGAGGTCACAGAGTAATTATGTC 851  
 Qy 270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIle 289  
 Db 852 CAAGAGTGGCGCGAATTCCTTCTCAGAGAGCGCGCATCCAGATGGAGATCATC 911  
 Qy 290 ValIysValAsnGlyArgProLeuValAspSerSerGluLeuGlnValIleValLeuThr 309  
 Db 912 GTCAAGGTCAAACGGGCGCTCTCTAGTGAAGTGAAGTGAAGTGAAGGCGTCTGACC 971  
 Qy 310 GlnSerProLeuLeuGluValArgArgGlyAsnAspLeuLeuPheSerIleAla 329  
 Db 972 GAGCT 1031  
 Qy 330 ProGluValValMet 334  
 Db 1032 CTTGAGGTGTCATG 1046

RESULT 3  
 LOCUS AY280665 2543 bp mRNA linear PRI 23-MAY-2003  
 DEFINITION Homo sapiens pregnancy-related serine protease HTRA3 mRNA, complete cds; alternatively spliced.  
 ACCESSION AY280665  
 VERSION AY280665.1 GI:31044217  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 2543)  
 Nie,G.Y., Hampton,A., Li,Y., Findlay,J.K. and Salomonson,L.A.  
 Identification and cloning of two isoforms of human high-temperature requirement factor A3 (HTRA3), characterization of its genomic structure and comparison of its tissue distribution with Htra1 and Htra2  
 JOURNAL Biochem. J. 371 (Pt 1), 39-48 (2003)  
 MEDLINE 22533278  
 PUBMED 12513693

REFERENCE 2 (bases 1 to 2543)  
 AUTHORS Nie,G., Hampton,A., Li,Y., Findlay,J.K. and Salomonson,L.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-APR-2003) Prince Henry's Institute of Medical Research, 246 Clayton Rd., Melbourne, Victoria 3168, Australia  
 FEATURES  
 source  
 1. 2543  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4p16.1"  
 /feature\_type="heart"  
 193..1554  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="pregnancy-related serine protease HTRA3"  
 /protein\_id="AAP42282.1"  
 /db\_xref="GI:31044218".  
 /translation="MOARALLAAALAAARPPAPPCARDVSRPCRCGYV  
 PDLCNCTLVASAESEPCGFLDPCGESIECYRGLCRCHSHAVCTDHTYANVCA  
 LQASRRALQSGTPVRQLQKACPLGLHQLSSPRKFNITAVETIAVAVHIELE  
 LRRLPLGRNVPLSSGGLFMSBAGLIITNNAVVSNSAARGQOLKVOLONGSYEAT  
 IKDIDKSDIATIKIHPKKLPLVILGSHADLPGEFVAIGSPFALONTVTGIVST  
 AOEGBELGLRSDMDYIOTDAILNYNSGGLPVNLDGEVIGINTLVKTGISPAIS  
 DRITRLTERODKQIDMKRFRIGIMRTTPTLNDKASNDPPEVSGIYQEV  
 PMSPSQRGIQDSDIIVKNGRPLVDSSEIQEAVLTESPLLEVRKNDLPLFAPE  
 VVN"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1 666-130 Length: 2543  
 Score: 1635.00 Matches: 325  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 97.38% Indels: 0  
 DB: 9 Gaps: 0  
 US-10-617-443b-2 (1-334) x AY280665 (1-2543)  
 Qy 10 GlyLeuHisGlnLeuSerSerProArgTYrLYsPheAsnPheIleAlaAspValIleGlu 29  
 Db 577 GGTCTCAGCAGCTGAGACAGCCCGCTCAAGATTCATCTTGTGTGTCAGTGTGAG 636  
 Qy 30 LysIleAlaProAlaValAlaHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49  
 Db 637 AAGATCGACACACCGGTGTCACATAGAGCTCTCTCTGAGACACCGCGTGTGGCGC 696  
 Qy 50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69  
 Db 697 AACGTGCCCTGTCCAGCGGTCTGCTTCATATGTCAGAGGCGCGCTGATCATCACC 756  
 Qy 70 AsnAlaHisValAlaSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuYsValGln 89  
 Db 757 AATGCCACCTGCTGTCTCAGCAACAGTCTGCGCCCGGAGGAGGACGATCAAGTGTCCAG 816  
 Qy 90 LeuGlnbngiYAspSerTYrGluAlaThrIleYsAspIleAspLYsSerAspIle 109  
 Db 817 CTACAGATGGGAGCTCTCTTAGAGCCACCATCAAGACATCGACAAAGAGTGGACATT 876  
 Qy 110 AlaThrIleYsIleHisProLYsLYsLeuProValLeuLeuGlyHisSerAla 129  
 Db 877 GCCACCATCAAGATCCCAAGAAAAGCTCCCTGTTGTTGCTGGTCACTCGGCC 936  
 Qy 130 AspLeuArgProGlyYgluPheValValAlaIleGlySerProPheAlaLeuGlnAspThr 149  
 Db 937 GACCTGGGCTGGGAGATTGTGTGGCATGCGCAGTCCCTTCCCTCAGAACACACA 966  
 Qy 150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAsp 169  
 Db 997 GTGACAAACGGGACATGTCAGCATGCGCAGGGAGGAGGAGGAGTGGGCTCCGGGAC 1056  
 Qy 170 SerAspMetAspTYrIleGlnThrAspAlaIleIleAsnTYrGlyAsnSerGlyGlyPro 189

```

Db      1057 TCGGACATGACATACATCCAGACGATGCAATCACTAGAGGAACTCCGGGGACCA 1116
Qy      190 LeuValAsnLeuAspGlyValValIleGlyIleAsnThrLeuValThrAlaGlyIle 209
Db      1117 CTGGTAACCTGGATGGAGGATCATTTGGCATCAACAGCTCAAGCTCAGCGCTGGCATC 1176
Qy      210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGluAspArgGln 229
Db      1177 TCTTTGGCATCCCTCCAGACCGCATCACAGGTTCTTCACAGAGTTCCAAAGACAGCAG 1236
Qy      230 IleValAspTrpIleValAspArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
Db      1237 ATCAAAAGACTGGAAAGAGCCCTTCATCGCATCGCATCGAGCATCAACCAAGCCCG 1296
Qy      250 ValAspGluLeuValAsaSerAspProAspPheProGluValSerSerGlyIleIleVal 269
Db      1297 GTGGATGAGCTGAAGGCCACCAACCGGACTTCCAGAGGTCAGCAGTGAATTATGTG 1356
Qy      270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleIle 289
Db      1357 CAAGAGGTTGGCGCGAATTCACCTTTCAGAGAGGCGGCATCCAAAGTGTGATCATCATC 1416
Qy      290 ValIleValAsnGlyValAspProLeuValAspSerSerGluLeuGlnGluValLeuThr 309
Db      1417 GTCAAGGTCAGGAGCGCTCTCTAGTGACTCGAGTGCAGGAGCGCGCTGTGACC 1476
Qy      310 GluSerProLeuLeuLeuGluValAlaArgGlyValAsnAspPheLeuPheSerIleAla 329
Db      1477 GAGCTCTCTCTCCACTGAGAGTGGCGGGGAAACGACCTCTCTTCAAGCATCGCA 1536
Qy      330 ProGluValAlaMet 334
Db      1537 CCTGAGGTGTCTATG 1551

RESULT 4
AY040094      2554 bp      mRNA      linear      PRI 05-JUN-2002
LOCUS      Home sapiens serine protease HTRA3 mRNA, complete cds.
DEFINITION      AY040094
ACCESSION      AY040094.2 GI:21326482
VERSION      1
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS      Southan,C. and Puntia,P.K.
TITLE      Direct Submision
JOURNAL      Submitted (14-JUN-2001) Computational Biology, Gemini Genomics, 162
Science Park, Milton Rd., Cambridge CB4 0GH, UK
2 (bases 1 to 2554)

REFERENCE
AUTHORS      Southan,C.
TITLE      Direct Submision
JOURNAL      Submitted (05-JUN-2002) Proteome Discovery, Oxford Glycosciences,
The Forum, 86 Milton Park, Abingdon, Oxon OX14 4RY, UK
Sequence update by submitter
REMARK      On Jun 5, 2002 this sequence version replaced gi:14700055.
COMMENT      location/Qualifiers
FEATURES
Source      1..2554
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /chromosome="4"
             /map="between D4S412 and D4S1601"
             /note="S1C serine peptidase family; similar to bacterial
             heat-shock protease HTRA"
             /codon_start=1
             /product="serine protease HTRA3"
             /protein_id="AAK1475.2"
             /db_xref="GI:21326483"
             /translation="MQARALLAALALALAREPPAAPCPARCDVSRCPSPCGYV

```

```

ORIGIN
Alignment Scores:
Pred. No.:      1,67e-130      Length:      2554
Score:      1635.00      Matches:      325
Percent Similarity:      100.00%      Conservative:      0
Best Local Similarity:      100.00%      Mismatches:      0
Query Match:      97.38%      Indels:      0
DB:      9      Gaps:      0

US-10-617-443b-2 (1-334) x AY040094 (1-2554)

Qy      10 GlyLeuHi sGlnLeuSerSerProArgTrpLysPheAsnPheIleAlaAspValAlaGlu 29
Db      606 GTCTCCACAGCTGAGCAGCAGCCCGCTTCAAGATTCACTTCACTTCTGACGTGGAGAG 665
Qy      30 LysIleAlaProAlaValAlaIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
Db      666 AAGATCCACAGCGCGTGTCCACATAGAGCTCTTCTCGAGACACCCGCTGTGTCCGC 725
Qy      50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
Db      726 AACGTGCCCTGTGTCAGGCGTGTGCTTCAATCATGTCAAGGCCCGGCTGATCATCACC 785
Qy      70 AsnIleHisValIleSerSerAsnSerAlaAlaProGlyValArgGlnLeuValGln 89
Db      786 AATGCCACGTGTGTGTCCAGACAGTGTGCCCCGGGAGGAGCGACGATCAAGTGCAG 845
Qy      90 LeuGlnAsnGlyAspSerTrpGlyAlaIleThrIleLysAspIleAspIleLysSerAspIle 109
Db      846 CTACAGAAATGGGGACTCTTATGAGCCACCATCAAGCATCGACCAAGAGTGGGACATT 905
Qy      110 AlaThrIleLysIleHisProLysLysLeuProValLeuLeuGluValHisSerAla 129
Db      906 GCCACATCAAGATCATCTCCAAAGAAAGCTCCCTGTGTGTGTGTGTGTGTGTGTGTGT 965
Qy      130 AspLeuArgProGlyGluPheValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
Db      966 GACCTGCGGCTGGGGAGATTGTGTGTGTCATGTGCGCATGTGCTTCCCTTACAGACACA 1025
Qy      150 ValThrThrGlyIleValSerThrAlaGlnArgGluArgGluLeuGlyLeuArgAsp 169
Db      1026 GTGACACAGGAGCATCTCTACGACTGCTCCAGCGGAGGAGGAGGAGCTGGGCTCCGGGAC 1085
Qy      170 SerAspMetAspTrpIleGlnThrAspAlaIleIleAsnTrpGlyValAsnSerGlyValPro 189
Db      1086 TCCGACATGAGACTACATCCAGACGAGTGCATCATCACTAGAGGAACTCCGGGGACCA 1145
Qy      190 LeuValAsnLeuAspGlyValValIleGlyIleAsnThrLeuValThrAlaGlyIle 209
Db      1146 CTGGTAACCTGGATGGAGGATCATTTGGCATCAACAGCTCAAGCTCAGCGCTGGCATC 1205
Qy      210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGluAspArgGln 229
Db      1206 TCTTTGGCATCCCTCCAGACCGCATCACAGGTTCTTCACAGAGTTCCAAAGACAGCAG 1265
Qy      230 IleValAspTrpIleValAspArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
Db      1266 ATCAAAAGACTGGAAAGAGCCCTTCATCGCATCGATCGAGCATCAACCAAGCCCG 1325
Qy      250 ValAspGluLeuValAsaSerAspProAspPheProGluValSerSerGlyIleIleVal 269
Db      1326 GTGGATGAGCTGAAGGCCACAGACCCGGACTTCCAGAGGTCAGTGTGAATTATGTG 1385
Qy      270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleIle 289

```









QY 109 IleaIaThrlIeYsIlleHlSpRoYslyblybLeuProValleuLeuGlyYHlSsr 128  
DB 826 ATTCGACCATATTGATATCCACCCCAAGAAAGCTCCCTGTGTGTCTGCTGGGTCTCA 885  
QY 129 AlaApLeuAArgProGlyGluPheValAlaIleGlySerProPheAlaLeuGlnAn 148  
DB 886 GCAAGACCTCGGCTGGCAGATTCGTGTGGCCATGCGAGCCCTTTCCTCGCAAGAC 945  
QY 149 ThrValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuArg 168  
DB 946 ACCGTGACAAAGGGCATTTGTACAGACTGCCAGCGGAGATGCAAGAGCTGGGTCTCCGG 1005  
QY 169 AapSerAapMeAapTrrIleGlnThrAapAlaIleIleAaNTrrGlyYbSserGly 188  
DB 1006 GACTCGACATGAGACTATATCCAGACCGATGCGCATCATATTAAGGAACTCAGAGGA 1065  
QY 189 ProLeuValAaLeuAapGlyGluValIleGlyIleAaNTrrLeuYsValThrAlaGly 208  
DB 1066 CCGCTGTGAACTGGATGGCAGATGCGATCGCATCAACGCTCAAGGTTGAGCTGGC 1125  
QY 209 IleSerPheAlaIleProSerAapArgIleThrArgPheLeuThrGluPheGlnApy 228  
DB 1126 ATCTCTTTGCCATCCCTCAAGTCGATCAACAGCTTCTCTGTGAGTTCCAAACAG 1185  
QY 229 GlnIleLyAapTrrIlyblybArgPheIleGlyIleArgMeAArgThrIleThrProSer 248  
DB 1186 CATGTGAAAGACTGGAAGAGCCCTTCATTTGCGATCGGATGCGAGCAATCAGCCAGG 1245  
QY 249 LeuValAapGluLeuYsAlaSerAaProAapPheProGluValSerSerGlyIleTyr 268  
DB 1246 TTGGTGGAGAACTGAAGCCGCCAACCAAGATTTTCAGCGGTCAAGTGAATATAT 1305  
QY 269 ValGlnGluValAlaProAaNTrrProSerGlnArgGlyIleGlnApyGlyAapIle 288  
DB 1306 GTTCAGAGAGTGGTCCCAATTCACCTTCAGAGAGAGGCAATCAAGATGCGCAGATC 1365  
QY 289 IleValIyValAaNTrrArgProLeuValAapSerSerGluLeuGlnGluValAaLeu 308  
DB 1366 ATCGTAAAGTCAATGAGCCGCCCTGCGGATTTCCAGAGCTGCAAGAGGAGATCCGG 1425  
QY 309 ThrGluSerProLeuLeuGluValArgArgGlyAaNTrrAapPheLeuPheSerIle 328  
DB 1426 AACGATCTTCACTCTGCTGGAGGTGCGGCGAGCAATGATCTCTCTTCAGCATC 1485  
QY 329 AlaProGluValAaMet 334  
DB 1486 ATCCCTGAGGTGTCATG 1503

RESULT 8  
LOCUS C0879976 1338 bp DNA linear PAT 11-OCT-2004  
DEFINITION Sequence 40 from Patent WO2004083241.  
ACCESSION C0879976  
VERSION C0879976.1 GI:54033832  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS Sakamoto, T. and Takeda, S.  
TITLE Btc-interacting proteins and use thereof  
JOURNAL Patent: WO 2004083241-A 40 30-SEP-2004;  
Takeda Chemical Industries, Ltd. (JP)  
LOCATION/Qualifiers  
FEATURES  
source 1..1338  
/organism="Mus musculus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:10090"

ORIGIN  
Alignment Scores:

Pred. No.: 2,43e-121 Length: 1338  
Score: 1524.00 Matches: 300  
Percent Similarity: 96.01% Conservative: 13  
Best Local Similarity: 92.02% Mismatches: 13  
Query Match: 90.77% Indels: 0  
Gaps: 0  
DB: 6  
US-10-617-443b-2 (1-334) x C0879976 (1-1338)

QY 9 AlaGlyLeuHlIleGlnLeuSerProArgTrrYsPheAaPheIleAlaAapValVal 28  
DB 358 TCTGTCTTCCACACAGCTGATCCGCGGTACAGTTCATGATGCGATGAGTGG 417  
QY 29 GlnYsIleAlaProAlaValAlaIleGlnLeuPheAaArgHlProLeuPheGly 48  
DB 418 GAGAAATTTGGCCAGCTGTGTCCACTTAAGCTTTTCTGAGACACCCCTGTTGGC 477  
QY 49 ArgAaNTrrProLeuSerSerGlySerGlyPheIleMeAArgGluAlaGlyLeuIle 68  
DB 478 CGGATGTGCGGCTGTCCAGTGGCTCGGGCTTCATCATGTGAGAAAGCCGTTGATCGTC 537  
QY 69 ThrAaNTrrAlaValAlaSerSerAaNTrrAlaProGlyAArgGlnLeuYsVal 88  
DB 538 ACCAAGCCCAAGTGTCTCCAGCTCCAGCACTGCTCCGCGGAGAGTGAAGTGG 597  
QY 89 GlnLeuGlnAaNTrrAapSerTrrGluAlaThrIleYsAapIleAapYsSerAap 108  
DB 598 CAGCTGCAAGATGGGATGCTCATAGGCCACATCCAGACATGCAAGAAATCGGAG 657  
QY 109 IleaIaThrlIeYsIlleHlSpRoYslyblybLeuProValleuLeuGlyYHlSsr 128  
DB 658 ATTCGACCATATTGATATCCACCCCAAGAAAGCTCCCTGTGTGTCTGCTGGGTCTCA 717  
QY 129 AlaApLeuAArgProGlyGluPheValAlaIleGlySerProPheAlaLeuGlnAa 148  
DB 718 GAGAGCTCGGCTGGGAGATTCGTGTGGCATGGAGCCCTTTCCTCGCAAGAC 777  
QY 149 ThrValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuArg 168  
DB 778 ACCGTGACAAAGGGCATTTGTACAGACTGCCAGCGGAGATGCAAGAGCTGGGTCTCCGG 837  
QY 169 AapSerAapMeAapTrrIleGlnThrAapAlaIleIleAaNTrrGlyYbSserGly 188  
DB 838 GACTCAGACATGAGCTATATCCAGACCGATGCCATCATCAATTAAGGAAACTCAGAGGA 897  
QY 189 ProLeuValAaLeuAapGlyGluValIleGlyIleAaNTrrLeuYsValThrAlaGly 208  
DB 898 CCGCTGTGAACTGGATGGCAGGTCATGCGCATGACGACGCTCAAGGTTGCAACTGGC 957  
QY 209 IleSerPheAlaIleProSerAapArgIleThrArgPheLeuThrGluPheGlnApy 228  
DB 958 ATCTCTTTGCCATCCCTCAAGTCGATCAACAGCTTCTCTTGAAGTTCAAAACAG 1017  
QY 229 GlnIleLyAapTrrIlyblybArgPheIleGlyIleArgMeAArgThrIleThrProSer 248  
DB 1018 CATGTGAAAGACTGGAAGAGCGCTTCATTTGCGATCGGATGCGAGCAATCAAGCAAT 1077  
QY 249 LeuValAapGluLeuYsAlaSerAaProAapPheProGluValSerSerGlyIleTyr 268  
DB 1078 TTGGTGGAGAACTGAAGCCGCCCAACCAAGCTTTCCAGCGGTCAAGATGATATAT 1137  
QY 269 ValGlnGluValAlaProAaNTrrProSerGlnArgGlyIleGlnApyGlyAapIle 288  
DB 1138 GTTCAGAGAGTGGTCCCAATTCACCTTCTCAGAGAGGAGCATCAAGATGGCGACATC 1197  
QY 289 IleValIyValAaNTrrArgProLeuValAapSerSerGluLeuGlnGluValAaLeu 308  
DB 1198 ATCGTCAAGTCAATGAGCCGCCCTGCGGATTCAGAGACTCAGAGAGCAAGCTCCGG 1257  
QY 309 ThrGluSerProLeuLeuGluValArgArgGlyAaNTrrAapPheLeuPheSerIle 328  
DB 1258 AACGATCTTCACTCTGCTGGAGGTGCGGCGAGCAATGATGATCTCTCTTCAGCATC 1317



TITLE  
JOURNAL  
COMMENT

Tickle, C. and Wilson, S.A.  
Direct Submission  
Submitted (09-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: chickesdbms.umist.ac.uk  
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA  
sequencing project.

This sequence is from the  
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
from a library constructed by Elizabeth Bosch. cDNA was prepared  
from RNA extracted from small intestine, normalised, and poly  
A-trimmed. EcoRI-NotI cut cDNA was then ligated into the vector.  
Vector: pBluescript II KS(+), Site\_1: EcoRI, Site\_2: NotI Host:  
Escherichia coli DH10B.

## FEATURES

## source

1. 2007  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Compton Line 151"  
/db\_xref="taxon:9031"  
/clone="CHEST625h5"  
/clone\_1lb="CSEQCIN56"  
/dev\_stage="adult"

## ORIGIN

## Alignment Scores:

Pred. No.: 8,31e-115 Length: 2007  
Score: 1450.50 Matches: 291  
Percent Similarity: 91.15% Conservative: 18  
Best Local Similarity: 85.84% Mismatches: 21  
Query Match: 86.39% Indels: 9  
DB: 5 Gaps: 1

US-10-617-443B-2 (1-334) x CR353995 (1-2007)

QY 5 LeuprolaAserAla-----GlyleuHsiGlnleuSer 15  
DB 175 CTCCCGCCCTGGCTCAGCTGCAAGAAAGACCTGGCAATCGGATCACTTCCAGC 234  
QY 16 SerProArGTYrlySephAenPheleAlaAspValValGluylleAlaPProAlaVal 35  
DB 235 AGTCGAAGATACAAATTCATTCATGAGATGTGTGAAAAGATGCACTGCAGTT 294  
QY 36 ValHleleGluLeuPheleuAArgHAsProleuPheGlyArgAenValProleuSerSer 55  
DB 295 GTGCACATTAACCTTCTCAGGACCCCTCTGTGTGGAATGTCACATTTCCAGT 354  
QY 56 GlySerGlyPheleleuMetSerGluAlaGlyleuIleleThrAenAlaHisValSer 75  
DB 355 GGATCTGGCTTATTATGTCGATTCTGTTATTGTGACCAATGACACGCTGCTCA 414  
QY 76 SerAenSerAlaAlaPProGlyArgGlnGlnleuylsValGlnleuGlnAenGlyAenSer 95  
DB 415 AGTACCAATCTATATCAAGGAGACACAGTTGAAAGCTACAGAAATGGAGATACA 474  
QY 96 TyrGluAlaThrIlelySAspIleAspIlelySAspIleAlaThrIlelySleHis 115  
DB 475 TATGACCAACCATCAAGACATTCAGCAAGAAATCTGACATTCGAATTAAGATCCAC 534  
QY 116 ProlylelySleuProValleuLeuGlyHisSerAlaAspLeuArgProGlyGlu 135  
DB 535 CTTAAGAAAAAATTAACCGTATTGTACTGGGCACTCTGATCTTAAGACCGAGGAA 594  
QY 136 PheValValAlaIleGlySerProPheAlaIleuGlnAenThrValThrThrGlyIleVal 155  
DB 595 TTGTACTGCAATTTGAAAGCCATTGCTCTACAGAACTGTGACAAACAGGATTTGTC 654  
QY 156 SerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAenSerAspMetAspTyrIle 175  
DB 655 AGCAGCTGCTAGCGAAGATGCAAGAAAGATTGCTTGGGGAGCTCGATATGATTAATT 714  
QY 176 GlnThrAspAlaIleIleAenTyrGlyAenSerGlyGlyProleuValAenLeuAspGly 195  
DB 715 CAGACAGATGCTATTATCAATTATGCAACTCTGAGAGACCTTAATTAACTGGATGGT 774

QY 196 GluValIleGlyIleAenThrleuylsValThrAlaGlyIleSerPheAlaIleProSer 215  
DB 775 GAAGGATGGAAATTAACCTTGAAGGTATACACTGGAATTTCTTTCGATTCCTCGC 834  
QY 216 AspArgIleThrArgPheLeuThrGluPheGlnAspIleGlnIlelySAspTyrlyS 235  
DB 835 GACCCCATTAATCAAGTCTCTCAAGAGTCGCTTGACAAACAAATTAAGATAGCAAAAAG 894  
QY 236 ArgPheIleGlyIleArgMetArgThrIleThrProSerleuValAspGluLeuylsAla 255  
DB 895 CGTTTCATTGGACATCCGAATGCTGACATTAACCTGCTTGAGGAAGATGGACAC 954  
QY 256 SerAenProAenPheProGluValSerSerGlyIleTyrValGlnGluValAlaProAen 275  
DB 955 AATAATGCTGATTTTCTGATGTCAAGAGTGAATTTTCATCAATGAAGTTCTTCAAT 1014  
QY 276 SerProSerGlnArgGlyIleGlnAspGlyAspIleIleValylsValAsnGlyArg 295  
DB 1015 TCACCTTCTCATAGAGAGATATCCAGATGAGATATCATCGTTAAAGTCAATGGCGCT 1074  
QY 296 ProleuValAspSerSerGluLeuGlnGluValleuThrGlnSerProleuLeu 315  
DB 1075 CTTTGAATGACTTCCAGTACCTTCAGAGCTGTGATGAACGAATCACCTTACTACTT 1134  
QY 316 GluValArgArgGlyAsnAspAspLeuPheSerIleAlaProGluValAlaMet 334  
DB 1135 GAACTTCGAGAGGAAATGATGACTTACTATTAACTTGAAGCTGGAATGTCTATG 1191

RESULT 11  
LOCUS BX929862 2176 bp mRNA linear VRT 02-FEB-2004  
DEFINITION Gallus gallus finished cDNA, clone CHEST656K24.  
ACCESSION BX929862  
VERSION BX929862.1 GI:41630390  
KEYWORDS  
SOURCE  
ORGANISM Gallus gallus (chicken)

REFERENCE  
AUTHORS Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E.,  
Croning, M.D.R., Davies, R.M., Francis, M.D., Graham, D.V.,  
Hubbard, S.J., Humphrey, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,  
Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,  
Tickle, C. and Wilson, S.A.

TITLE  
JOURNAL  
COMMENT Direct Submission  
Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: chickesdbms.umist.ac.uk  
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA  
sequencing project.

This sequence is from the  
BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
from a library constructed by Elizabeth Bosch. cDNA was prepared  
from RNA extracted from limbs, normalised, and poly A-trimmed.  
EcoRI-NotI cut cDNA was then ligated into the vector. Vector:  
pBluescript II KS(+), Site\_1: EcoRI, Site\_2: NotI Host: Escherichia  
coli DH10B.

## FEATURES

## source

1. 2176  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, HiseX"  
/db\_xref="taxon:9031"  
/clone="CHEST656K24"  
/clone\_1lb="CSEQCIN38"  
/dev\_stage="stage 36"

## ORIGIN

## Alignment Scores:

Pred. No.: 9,23e-115 Length: 2176  
Score: 1450.50 Matches: 291

Percent Similarity: 91.15%  
 Best Local Similarity: 85.84%  
 Query Match: 86.39%  
 DB: 5  
 Gaps: 1

US-10-617-443B-2 (1-334) x BX929862 (1-2176)

Conservative: 18  
 Mismatches: 21  
 Indels: 9

5 Leuproalaserla-----GlyLeuHieGlnLeuSer 15  
 329 CTGCCCCCGCTGCTCAGCTGCAAGAACGCTCCGATCGGCTCACTTACGCTCCAGC 388  
 16 SerProArgTyrIlyPheAsnPhelIleAlaSpValValGluYerIleAlaProAlaVal 35  
 389 AGTCACAGATACCAATTCATCTCATAGACATGATGGGAAAGATTGCACTGCACTT 448  
 36 ValHieIleGluLeuPheLeuHieAspProLeuPheGlyAgaAsnValProLeuSer 55  
 449 GTGCACATTGAACCTTCTCAGGACCCCTGTTGGTCGAAATGTCCTCCACTTCCAGT 508  
 56 GlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThrAsnAlaHieValValSer 75  
 509 GGATCTGGCTTATATATCTGATCTGCTGTTTAAATGACCAATGACACAGTGGTCA 568  
 76 SerAsnSerAlaAlaProGlyArgGlnGlnLeuYsValGlnLeuGlnAsnGlyAspSer 95  
 569 AGTACCAATGCTATATACGAGGACACAGTGAAGTCCAGCTACAGATGAGATACA 628  
 96 TyrGluAlaThrIleIlyAspIleAspIlySerAspIleAlaThrIleYerIleHie 115  
 629 TATGACGACACCAATGACGATTCACAAATCTGACATTCGACAAACAAATGAATCCAC 688  
 116 ProIlyIlyIlyLeuProValLeuLeuGlnIlyHieSerAlaAspLeuArgProGlyGlu 135  
 689 CTTAGAAATTAATTAACCGTATGTTACTGGGACCTGCTGATCTTACAGCCAGGAGAA 748  
 136 PheValValAlaIleGlySerProPheAlaLeuGlnAsnThrValThrThGlyIleVal 155  
 749 TTTGTAGGCAATTGGAAGCCATTGGCTTACGAAACATGTCGACACAGATTTGTC 808  
 156 SerThrIleGlnIlyArgGluGlnLeuGlnLeuArgAspSerAspMetAspTyrIle 175  
 809 AGCATCTCTCAGCAGATGCGAAAGAACTTGGCTGGGCGCTCGGATATGATTAACATT 868  
 176 GlnThrAspAlaIleIleAsnThrGlyAsnSerGlyGlyProLeuValAsnLeuAspGly 195  
 869 CAGGACATGCTATATCAATTAATGAGCACTGAGAGACCTTGAATTAATCTGAGTGT 928  
 196 GluValIleGlyIleAsnThrLeuIlyValThrAlaGlyIleSerPheAlaIleProSer 215  
 929 GAATGATTTGGAATTAACCTTGAAGTTTACAGCTGGAATTTCTTTCATTTCCGTCG 988  
 216 AspArgIleThrArgPheLeuThrGluPheGlnAspIlyGlnIleYsAspTyrIlyAs 235  
 989 GACCGCATTTACTCAGTCTCTCAGAGAGCGCTTGAACAAACAAATTAAGATGACAAAAG 1048  
 236 ArgPheIleGlyIleArgMetArgThrIleThrProSerLeuValAspGluLeuIlyAla 255  
 1049 CGTTTACTGCAATCCGAAATGCTGACAAATACCTCTCTGTTGGAGAAATTAACACAC 1108  
 256 SerAsnProAspPheProGluValAsnSerSerGlyIleTyrValGlnGluValAlaProAsn 275  
 1109 AATTAATCTGATTTCTCTGATGTCAGAGTGGAAATTTTACATGACATGATTTCCAAAT 1168  
 276 SerProSerGlnArgGlyIleGlnAspIlyAspIleIleValIlyValAsnGlyArg 295  
 1169 TCACCTTCTCATAGAGAGGATATCCAGATGAGATATCATCGTTAAAGTCAATGGCGGT 1228  
 296 ProLeuValAspSerSerGluLeuGlnGluValAlaLeuThrGlnSerProLeuLeuLeu 315  
 1229 CCTTGTAGCTTCTCAGTGAATTCGAGAGGCTGTGATGAACGATACCTTACTTACTT 1288  
 316 GluValArgArgGlyAsnAspAspLeuPheSerIleAlaProGluValIleMet 334

DB 1289 GAACTGCGAGAGGAATGATGACTTACTATTATTAACATGAGCCCTGAATGTCATG 1345

RESULT 12  
 EC078402  
 LOCUS  
 DEFINITION  
 accession  
 version  
 keywords  
 ORGANSIM  
 SOURCE

EC078402 2080 bp mRNA linear VRT 16-AUG-2004  
 Danio rerio zgc:91963, mRNA (CDNA clone MGC:91963 IMAGE:7041654),  
 complete cde.  
 EC078402  
 MGC.  
 Danio rerio (zebrafish)  
 Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 2080)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Shapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Muihah, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalski, U., Smallos, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 2080)  
 Director MGC Project.  
 Direct Submission  
 Submitted (19-JUL-2004) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2530,  
 USA

REMARK  
 COMMENT  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Len Zon, Harvard  
 cDNA Library Preparation: Open Biosystems  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc.mgc@nrl.nih.gov](mailto:nisc.mgc@nrl.nih.gov)  
 Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Heghighi, P.,  
 Hansen, N., Ho, S.-L., Karlins, B., Kong, P., Larc, P., Legaspi, R.,  
 McDowell, J., Pearson, C., Stancitop, S., Thomas, P.J., Touchman, J.W.,  
 Tsurgon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAP Plate: 178 Row: 1 Column: 24  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis, similarity but not identity to protein.  
 Location/Qualifiers  
 1. 2080

```
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="MGC:91963 IMAGE:7041654"
/tissue_type="Whole, adult zebrafish"
/clone_id="N1H_ZGC_10"
/lab_host="DH10B"
/notes="Vector: pExpress1"
1..2080
/gene="zgc:91963"
/db_xref="LOCUSID:445108"
143..1612
/gene="zgc:91963"
/codon_start=1
/product="zgc:91963"
/protein_id="AAH78402.1"
/db_xref="GI:50418457"
/db_xref="LOCUSID:445108"
/translation="MOVILLAATFTFDLAIABLKNSPKCDVSKCPSPCSGY
VPRNCCLVCSLGEAPGGRKEDVLCGDLCEKPTGRKLSKGVCCCTHRCVGTD
GKTYGVNCKRRTSRKAQORLSVITQHKHSCAPSSPAHNSPRYFNIADVE
KIAVAVHVELFLNPLRGRHVPVLSGSGPTMOSGLITVNAHVASATVGRHLR
VQADGQTYEASIRIDDKSDIATIKNPKKLQVSLRSADLRPEFVAVIGSFA
LQNTVTGIVSTIRDRKELGIRDSDMGYIQTDAITNGSGPLVNLDEGVTGINTL
KVTAGISFALPSDRINKFLDESNDKQKQKRVNRVNTYQSOAMRTASDVNPMKRFI
GIKMWLTLENLVHELKWNHNPAPDIDSGILVHEVADSPAOKXGLESDDIYKLNQHP
LMTGELDEALIQVMDPLLEVRGNDDLFPNBPQILMQ"
```

ORIGIN

Alignment Scores:  
Prid. No.: 1,36e-96 Length: 2080  
Score: 1238.50 Matches: 245  
Percent Similarity: 81.82% Conservative: 43  
Best Local Similarity: 69.60% Mismatches: 41  
Query Match: 73.76% Indels: 23  
DB: Gaps: 1

US-10-617-443B-2 (1-334) x BC078402 (1-2080)

```
QY 6 ProLaserAlaglyLeuHieGlnLeuSerProAgtYrlyrPheanpHeila 25
    |||:::|||||
DB 551 CCTCAACTCAGATCCTCGCCATCTGAACGCCGAGTCAAGTTTAACTTCATCGCT 610
QY 26 AspValValGluYrYrValProAlaValHieGlnLeuPheLeuAArgHiePro 45
    |||:::|||||
DB 611 GATGGGGGAGAGATCGCCCTGCTGTGTGTCACGTCGAGCTCTTCTCAACATCA 670
QY 46 LeupheGlyAgtAanValProLeuSerSerGlySerGlyPheileMetSerGluAlaGly 65
    |||:::|||||
DB 671 CTCTTGGCCGTCAGTACCTTGTTCAGCGGCTCTGATTCATCATGACCACATCGCT 730
QY 66 LeuileileThraValAhiEvaValSerSerAanSerAlaValProGlyAArgGln 85
    |||:::|||||
DB 731 CTGATTGGACCAAGCTCATGTGTTGTCAGTACGACAGTAAACAGGCGCCAGCAT 790
QY 86 LeuYrValGlnLeuGlnAanglyAspSerYrGluAlaThrIleLeuAspIleAspYr 105
    |||:::|||||
DB 791 CTACAGAGTCAGCTCACAGTGGCCAAACCTACGAGGCTTCATGAGACATTCACAG 850
QY 106 LysSerAspIleAlaThrIleYrIleHieAspProYrYrLeuProValLeuLeu 125
    |||:::|||||
DB 851 AATATCGACATCGCCACATCAAGATTAAACCAAGAAAGAAATTACAGGCTTATCATTTA 910
QY 126 GlyHieSerAlaAspLeuAArgProGlyGluPheValAlaAlaIleGlySerProPheAla 145
    |||:::|||||
DB 911 GGTGGCTGTGTCGACCTGAGGCTGTGTGTTGTGTTGTCATCGGCGCCCTTTGTCT 970
QY 146 LeuGlnAanThrValThrGlyIleValSerThrAlaGlnAArgGluGlyAArgGluLeu 165
    |||:::|||||
DB 971 CTTCAACAACAGTCAAGACGATCGTCAGCAACAACAAGAGATGGAAGAAGGCTC 1030
QY 166 GlyLeuAArgAspSerAspMetAspYrIleGlnThrAspAlaIleIleAanYrGlyAan 185
    |||:::|||||
```

```
DB 1031 GGTATCCGGAGCTCTGACATGGGCTACATCCAGACCCGATCCATCATTTATGGAAC 1090
QY 186 SerGlyYrProLeuValAanLeuAanGlyGluValIleGlyIleAanThrLeuYrVal 205
DB 1091 TCAGAGAGTCCACATCGTTATTCGATGAGTGAAGTATGGAATTAACATTAAGAGTG 1150
QY 206 ThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrAArgPheLeuThrGluPhe 225
DB 1151 ACTGACAGGATCTCTTTTGGCATTCCTCCATCCAGACGATCAACAAGTTCTGTATGATCAT 1210
QY 226 GlnAspYrGlnIleYrAspYrPheYr----- 234
DB 1211 AATGACAAACGACGAAAGTCAAGCAAGGAGTATGACGACAAACTACAGCAGTCTCAA 1270
QY 235 -----LysAArgPheIleGlyIleAArgMet 242
DB 1271 GCTATGAGAGCTGCTCAGACGTAATGTCCTCCATGAAAGGTTTATGGAATCAAAATG 1330
QY 243 AArgThrIleThrProSerLeuValAspGluLeuYrValSerAanProAspPheProGlu 262
DB 1331 GTCACTTTAACAGAAACCTAGTCCATGAAATTGGAATGGCATATCCAGCCTTCCGTAT 1390
QY 263 ValSerSerGlyIleYrValGlnGluValAlaProAanSerProSerGlnAArgGlyYr 282
DB 1391 ATTGAAGTGGGATTTCTTGTTCATGAAGTTATCGCAGACTCCCTGCTCAGAAAGGTCGG 1450
QY 283 IleglnAspGlyAanPheIleValYrValAanGlyYrAArgProLeuValAspSerSerGlu 302
DB 1451 CTGAAATCAGGATGATTTATGATGAACTAAACGCCCATTCATTAAAGAACCTGAGAA 1510
QY 303 LeuGlnGluAlaValLeuThrGluSerProLeuLeuGluValAlaAArgGlyYrAspAsp 322
DB 1511 CTGACGAGAAAGCATTCAGATGACATGCCCTCTTTTGAGAGTTGCGCGGATTAATGAC 1570
QY 323 AspLeuLeuPheSerIleAlaProGluValAlaMet 334
DB 1571 GATCTGCTTTTAACTTGAACCTCAATCTCTCATG 1606
```

RESULT 13

AY280666 1953 bp mRNA linear PRI 23-MAY-2003  
LOCUS Homo sapiens pregnancy-related serine protease HTRA3 mRNA, complete  
DEFINITION cds; alternatively spliced.

ACCESSION AY280666  
VERSION AY280666.1 GI:31044219

SOURCE Homo sapiens (human)  
ORGANISM

REFERENCE 1 (bases 1 to 1953)  
AUTHORS Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.  
TITLE Nucleotide sequence and cloning of two isoforms of human

high-temperature requirement factor A3 (HTRA3), characterization of  
its genomic structure and comparison of its tissue distribution  
with HTRA1 and HTRA2

JOURNAL Biochem. J. 371 (Pt 1), 39-48 (2003)

MEDLINE 22533278  
PUBMED 12513693

REFERENCE 2 (bases 1 to 1953)  
AUTHORS Nie, G., Hampton, A., Li, Y., Findlay, J.K. and Salamonson, L.A.

TITLE Direct Submission  
JOURNAL Submitted (22-APR-2003) Prince Henry's Institute of Medical  
Research, 246 Clayton Rd., Melbourne, Victoria 3168, Australia

FEATURES location/Qualifiers  
1..1953

SOURCE /organism="Homo sapiens"  
/mol\_type="mRNA"

DB /db\_xref="taxon:9606"  
/chromosome="4"

CDS /map="4p16.1"  
/feature="heart"

CDS

108..1181



```

/note="alternatively spliced"
/codon_start=1
/product="pregnancy-related serine protease HTRA3"
/protein_id="AAP42283.1"
/db_xref="GI:31044220"
/translation="MQARALULALALALAREPPAPACPARCDVSRCPSPRCGGY
PDLNCCCLVCAASBGPBGPLDSPGSESLCIVGLCRSHAVCGTDHTVANVCA
LOASRALQSGTPVROLQACPLGLHASPXYKFNIPADIVKIAPAVHIEP
LRHPLFGSNVPLSSGSGFIMSEAGLITNAHVSNAPAPROOLKXUOLONGDSYEAT
IKDIDKSDIATIKHPKKKLPULLIGHNSADLRGEFVVALGSPFALONTTGTIVST
AQRERELGLRDSMDYITQDAIITNGNSGVLVNLDSVIGINTLKVTAGISPAIDS
DRITRFLTEFDKQIKAPSLAVH"

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 3,34e-86 Length: 1953
Score: 1117.00 Matches: 222
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.53% Indels: 0
DB: 9 Gaps: 0

```

US-10-617-443B-2 (1-334) x AY280666 (1-1953)

```

QY 10 G1yeuHhagInLeuSerSerProArgTyrLySPhenPheh1ealAaspValValGlu 29
Db 492 GGTCTCCACCGAGCTGAGACGCCCGCTACAGTTCACTTCTGAGCTGTGGAG 551
QY 30 Lye1leAlaProAlaValAlaHis1leGluLeuPheLeuArgHisProLeuPheG1yArg 49
Db 552 AAGATCCACACGACGCTGTGCACATAGAGCTTCTCGAGACACCCGCTTTGGCCGC 611
QY 50 AsnValProLeuSerSerGlySerGlyPhe1leMetSerGlyValAG1yLeu1le1leThr 69
Db 612 AACGTCGCCCTGTCCAGCGGTTGTGCTTCATCTGTCAGAGCGCGCTGATCATCC 671
QY 70 AsnAlaHisVal1SerSerAsnSerAla1a1aProG1yArgG1ngInLeu1yValG1n 89
Db 672 AATGCCACGCTGTCTCCAGCAAGTCTCCCGGCGAGGACGACCTCAAGTGCAG 731
QY 90 LeuG1naEng1yAapSerTyrGluAlaThr1le1yAap1leAap1y1y1ySerAap1le 109
Db 732 CTACAGATGGGAGCTCTCTATGAGGCCACCATCAAGACATCGAAGAGTCCGACAT 791
QY 110 AlaThr1le1y1leHisProLy1y1yLeuProVal1leuLeuEng1yHisSerAla 129
Db 792 GCCACCATCAAGATCCATCCAGAAAAAGCTCTGTGTGTGTGTGTGTGTGTGTGTGT 851
QY 130 AspLeuArgProG1yGluPheValValAla1leG1ySerProPheAlaLeuG1naSnThr 149
Db 852 GACCTGGCGCTGGGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 911
QY 150 ValThrThrG1y1leVal1SerThrAlaG1naG1ng1yArgG1uLeuG1yLeuArgAap 169
Db 912 GTGACAAAGGGCATCTGTCAGACATGCCAGCGGAGGAGGAGGAGCTGGGCTCGGGAC 971
QY 170 SerAspMetAapTyr1leG1nThrAapAla1le1leAanTyrG1yAanSerG1yPro 189
Db 972 TCCACATGAGCTACATCCAGACGAGATGCCATCTCACTACGGGAATCTCCGGGAGCA 1031
QY 190 LeuValAsnLeuAapG1yGluVal1leG1y1leAanThrLeu1yVal1ThrAlaG1y1le 209
Db 1032 CTGGTGAACCTGATGGGAGGATGTCATGCAACGCTCAAGGTCACGGCTGGGATC 1091
QY 210 SerPheAla1leProSerAapArg1leThrArgPheLeuThrG1uPheG1naAap1yG1n 229
Db 1092 TCCCTTGCATCCCTCAGACGCGATCAACAGTTCCTCAAGAGTTCCAAAGCAACAG 1151
QY 230 1le1y1y 231
Db 1152 ATCAAA 1157

```

RESULT 14

## ORIGIN

## Alignment Scores:

```

Pred. No.: 1.91e-85 Length: 819
Score: 1102.50 Matches: 235
Percent Similarity: 73.23% Conservative: 3
Best Local Similarity: 72.31% Mismatches: 10
Query Match: 65.66% Indels: 77
DB: 6 Gaps: 3

```

US-10-617-443B-2 (1-334) x CQ716317 (1-819)

```

QY 10 G1yeuHhagInLeuSerSerProArgTyrLySPhenPheh1ealAaspValValGlu 29
Db 73 GGTCTCCACCGAGCTGAGACGCCCGCTACAGTTCACTTCTGAGCTGTGGAG 132
QY 30 Lye1leAlaProAlaValAlaHis1leGluLeuPheLeuArgHisProLeuPheG1yArg 49
Db 133 AAGATCCACACGACGCTGTGCACATAGAGCTTCTCGAGACACCCGCTTTGGCCGC 192
QY 50 AsnValProLeuSerSerGlySerGlyPhe1leMetSerGlyValAG1yLeu1le1leThr 69
Db 193 AACGTCGCCCTGTCCAGCGGTTGTGCTTCATCTGTCAGAGCGCGCTGATCATCC 252
QY 70 AsnAlaHisVal1SerSerAsnSerAla1a1aProG1yArgG1ngInLeu1yValG1n 89
Db 253 AATGCCACGCTGTCTCCAGCAAGTCTCCCGGCGAGGACGACGCTCAAGTGCAG 312
QY 90 LeuG1naEng1yAapSerTyrGluAlaThr1le1yAap1leAap1y1y1ySerAap1le 109
Db 313 CTACAGATGGGAGCTCTCTATGAGGCCACCATCAAGACATCGAAGAGTCCGACAT 372
QY 110 AlaThr1le1y1leHisProLy1y1yLeuProVal1leuLeuEng1yHisSerAla 129
Db 373 GCCACCATCAAGATCCATCCAGAAAAAGCTCTGTGTGTGTGTGTGTGTGTGTGTGT 432
QY 130 AspLeuArgProG1yGluPheValValAla1leG1ySerProPheAlaLeuG1naSnThr 149
Db 433 GACCTGGCGCTGGGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 492
QY 150 ValThrThrG1y1leVal1SerThrAlaG1naG1ng1yArgG1uLeuG1yLeuArgAap 169
Db 493 GTGACAAAGGGCATCTGTCAGACATGCCAGCGGAGGAGGAGGAGCTGGGCTCGGGAC 552
QY 170 SerAspMetAapTyr1leG1nThrAapAla1le1leAanTyrG1yAanSerG1yPro 189
Db 553 TCCGACATGAGCTACATCCAGACGAGATGCCATCTCACTGAGT----- 597
QY 190 LeuValAsnLeuAapG1yGluVal1leG1y1leAanThrLeu1yVal1ThrAlaG1y1le 209
Db 597 ----- 597

```

CQ716317  
LOCUS CQ716317 819 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 2251 from Patent WO02068579.  
ACCESSION CQ716317  
VERSION CQ716317.1 GI:42277174  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homosapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.  
Kites, such as nucleic acid arrays, comprising a majority of  
humanexons or transcripts, for detecting expression and other uses  
thereof  
Patent: WO 02068579-A 2251 06-SEP-2002;  
PE Corporation (NY) (US)  
Location/Qualifiers  
1. 819  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## FEATURES

source

QY 210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229  
 Db 598 -----CCCAAGGACAGAGAGCCGGGACACCTG----- 624  
 QY 230 IleLysAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerIleu 249  
 Db 625 -----CCATCTGTG 633  
 QY 250 ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleIerVal 269  
 Db 633 ----- 633  
 QY 270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleIle 289  
 Db 634 -----CCCAAGACCTCACCTCGAGAGCGCGCATCCAGATGTGTACATCAGC 681  
 QY 290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnLysAlaValLeuThr 309  
 Db 682 GTCAAGGTCAACGGCGCTCTCTAGTGAAGTCAAGTGAAGTGAAGGCGGCTGACCC 741  
 QY 310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329  
 Db 742 GAGTCTCTCTCTCTCACTGAGAGGTGCGCGGGGAGAACGACGACCTCTTCAAGATCGCA 801  
 QY 330 ProGluValValMet 334  
 Db 802 CTGAGGTGTGATG 816  
 RESULT 15  
 LOCUS AY280664 1897 bp mRNA linear ROD 01-MAY-2004  
 DEFINITION Mus musculus pregnancy-related serine protease mRNA, complete cds;  
 accession AY280664  
 VERSION AY280664.1 GI:33358216  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eumetazoa; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS 1 (bases 1 to 1897)  
 TITLE Nie, G., Li, Y., Findlay, J. K. and Selamonsen, L. A.  
 IDENTIFICATION and cloning of a novel pregnancy-related  
 serine-protease (PRSP)  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1897)  
 AUTHORS Nie, G., Li, Y., Findlay, J. K. and Selamonsen, L. A.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-APR-2003) Prince Henry's Institute of Medical  
 Research, 246 Clayton Rd, Melbourne, Victoria 3168, Australia  
 location/Qualifiers  
 1. 1897  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="Swiss"  
 /db\_xref="taxon:10090"  
 /chromosome="5"  
 /tissue\_type="uterus"  
 127..1218  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="pregnancy-related serine protease"  
 /protein\_id="AA016593.1"  
 /db\_xref="GI:33358217"  
 /translation="MOARALLPATLALATLVALALARPAPAPCAPARQVSRCPSPR  
 CPGGYVPLDLCNCLVCAASBEGPCGRPLDSLCYRAGCRCHWTTCGTGHT  
 VADYCALOASRRALQVSGTPVQLQKACPSGLHQDTSPRYKFNITADVERKIAAY  
 VHIETPLRHPLFGNNVPLSSGSGFIMSEAGLIYVNAHVSSSTASGRQDLKVLQNG  
 DAYEATITODIDKSDIATIVIHKKKLPVLLGHSADLRPGSEVVAIGSPFALQNTVT  
 TGIIVTAQRDKELGLRSDPMYTOTDAIINYGSGGPLVNLDEVAIGINTLKVAAIGI  
 SPALPSDITRTFSEFQKVKALSPALH"  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4,536-81 Length: 1897  
 Score: 1057.00 Matches: 207  
 Percent Similarity: 97.76% Conservative: 11  
 Best Local Similarity: 92.83% Mismatches: 5  
 Query Match: 62.95% Indels: 0  
 DB: 10 Gaps: 0  
 US-10-617-443b-2 (1-334) x AY280664 (1-1897)  
 QY 9 AlaGlyLeuNHISglLeuSerSerProArgTyrlAspAsnPhelIleAlaAspValVal 28  
 Db 526 TCTGTCTCCACCACTGACACGCTCGGGTCAAGTTCAATTCATCGCCGATGTGTG 585  
 QY 29 GluLysIleAlaProAlaValAlaHisIleGluLeuPheLeuArgHisProLeuPheGly 48  
 Db 586 GAGAAATTCGCCACCTGTGTGTCACATAGAGCTTCTTCTAGACACCCCTGTTGGC 645  
 QY 49 ArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIle 68  
 Db 646 CGGAATGTCCGCTGTTCATAGTGGCTCGGCTTCATCATGTCAAGAACCCGTTGTATCGTC 705  
 QY 69 ThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysVal 88  
 Db 706 ACCAAGCCCAACGTCCTCAGCTCCAGCTCCAGCATGCTCCGCGGAGCAGCTGAAGTG 765  
 QY 89 GlnLeuGlnAsnGlyAspSerTyrlGluAlaThrIleLysAspIleAspLysSerAsp 108  
 Db 766 CAGCTGCAGATGGAGAGAGTGGCTATGAGCCACCATCCAGACATCCAGAAAGTCGGAC 825  
 QY 109 IleAlaThrIleLysIleHisProLysLysLysLeuProValLeuLeuLeuGlyHisSer 128  
 Db 826 ATTGCCAGATGTGAATCCACCCAGAAAAGCTCTGTGTGTGCTGCTGGCTACTCA 885  
 QY 129 AlaAspLeuArgProGlyGluPheValAlaIleGlySerProPheAlaLeuGlnAsn 148  
 Db 886 GCAGACCTGGCGCTCGGGAGAGTCTGTGTGGCCATCGGAGGCCCTTGTGCTGCAGAAC 945  
 QY 149 ThrValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuArg 168  
 Db 946 ACCGTGACACAGCGGATGTGACGCTCCAGCGGGATGGCAAGAGAGCTGGATCTCCGG 1005  
 QY 169 AspSerAspMetAspTyrlIleGlnThrAspAlaIleIleAsnTyrlGlyAsnSerGlyGly 188  
 Db 1006 GACTCAGACATGACTATATCCAGACCATGCTCATCATTTACGGGAATCAGAGAGA 1065  
 QY 189 ProLeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGly 208  
 Db 1066 CCCCTGTGTAACCTGGATGGCAGAGTCATCGGCATCAACAGCTCAAGGTTCAGCTGGC 1125  
 QY 209 IleSerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLys 228  
 Db 1126 ATCTCTTGTGCCATTCCTCCATGATGCATCAACGCTTCTCTGAGAGTCCAAAACAAG 1185  
 QY 229 GlnIleLys 231  
 Db 1186 CATGTGAAA 1194  
 Search completed: June 29, 2005, 23:48:50  
 Job time : 7535.54 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 29, 2005, 18:37:40 ; Search time 937.732 Seconds

(without alignments)  
2108.485 Million cell updates/sec

Title: US-10-617-443B-2  
Perfect score: 1679  
Sequence: 1 MHLALPASAGLHQLSPRYK.....LEVRGRNDLLFSLAPEVVM 334

Scoring table:  
BIOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h  
-Q/cgnt\_1/USPTO.spool/US10617443/runat\_29062005\_135931\_26052/app\_query.fasta\_1.718  
-DB=N\_Geneseq\_16Dec04 -QMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MTN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=p2n -NOM=ext -HEAPSIZE=500 -MTLEN=0 -MAXLEN=2000000000  
-USER=US10617443 @CGN\_1.1 586 @runat\_29062005\_135931\_26052 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEDBTRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_16Dec04:

1: geneeqn1980s:\*  
2: geneeqn1990s:\*  
3: geneeqn2000s:\*  
4: geneeqn2001as:\*  
5: geneeqn2001bs:\*  
6: geneeqn2002as:\*  
7: geneeqn2002bs:\*  
8: geneeqn2003as:\*  
9: geneeqn2003bs:\*  
10: geneeqn2003cs:\*  
11: geneeqn2003ds:\*  
12: geneeqn2004as:\*  
13: geneeqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	DB ID	Description
1	1679	100.0	3006	ADJ11355 Human PRS
2	1635	97.4	1044	Ad44189 Mouse TAN
3	1635	97.4	1316	AD001101 Human ser
4	1635	97.4	1319	AD001102 Human ser
5	1635	97.4	1359	AD001091 Human ser

6	1635	97.4	1375	12	AD001099	Ad001099 Human ser
7	1635	97.4	1378	12	AD001093	Ad001093 Human ser
8	1635	97.4	1563	3	AA57361	AA57361 cDNA enco
9	1635	97.4	1583	10	ACF16993	ACF16993 Human NOV
10	1635	97.4	1797	6	AA023855	AA023855 Human pro
11	1635	97.4	2040	3	AA523622	AA523622 NSBQ gene
12	1635	97.4	2040	6	AA142458	AA142458 Human mat
13	1635	97.4	2094	6	ABK51497	ABK51497 cDNA enco
14	1635	97.4	2543	8	ACF12810	ACF12810 Human PRS
15	1635	97.4	2554	12	AD001092	Ad001092 Human ser
16	1635	97.4	2577	3	AA57359	AA57359 cDNA enco
17	1632	97.2	1375	12	AD001106	Ad001106 Human ser
18	1632	97.2	1378	12	AD001105	Ad001105 Human ser
19	1608	95.8	960	12	AD001116	Ad001116 Human ser
20	1536	91.5	2450	8	ACF12809	ACF12809 Full leng
21	1524	90.8	1338	13	AD149833	Ad149833 Murine TA
22	1378	82.1	1868	4	AA541158	AA541158 cDNA enco
23	1378	82.1	1868	4	AA526920	AA526920 Human CDN
24	1144	68.1	1075	12	AD001100	Ad001100 Human ser
25	1122	66.8	1695	4	AA514885	AA514885 Human CDN
26	1117	66.5	1953	8	ACF12811	ACF12811 Human PRS
27	1105	65.8	1788	4	AA526848	AA526848 Human CDN
28	1057	63.0	1897	8	ACF12815	ACF12815 Full leng
29	1034.5	61.6	1185	3	AA229195	AA229195 cDNA enco
30	1034.5	61.6	1440	6	AA044188	AA044188 Human TAN
31	1034.5	61.6	1443	3	AA229176	AA229176 Recombina
32	1034.5	61.6	1443	10	AD113864	Ad113864 Osteoarth
33	1034.5	61.6	1443	10	AD113865	Ad113865 Osteoarth
34	1034.5	61.6	1894	10	AD062641	Ad062641 Human CDN
35	1034.5	61.6	1972	13	AD510059	Ad510059 Human the
36	1034.5	61.6	2034	7	AD573135	Ad573135 Human kid
37	1034.5	61.6	2036	6	ABV78051	ABV78051 Hypoxia-r
38	1034.5	61.6	2036	13	ADR42954	ADR42954 Breast ca
39	1034.5	61.6	2036	13	ADR40166	ADR40166 Human ser
40	1034.5	61.6	2039	10	AD113867	Ad113867 Osteoarth
41	1034.5	61.6	2054	10	ADJ56226	AdJ56226 Rat cDNA
42	1034.5	61.6	2054	10	AD113863	Ad113863 Osteoarth
43	1034.5	61.6	2075	2	AA754444	AA754444 cDNA enco
44	1034.5	61.6	2205	2	AAV04680	AAV04680 Human pre
45	1034.5	61.6	2437	11	ACN89676	ACN89676 Breast ca

## ALIGNMENTS

RESULT 1	
ID	ADJ11355 standard; cDNA; 3006 BP.
XX	
AC	ADJ11355;
XX	
DT	20-MAY-2004 (first entry)
XX	
DE	Human PRSS11-L cDNA encoding a serine protease Segid 1.
XX	
KW	human; PRSS11-L; gene; ss; S2 serine protease; S2/HTRA;
KW	cellular physiology; apoptosis; gene therapy; osteoarthritis; stress;
KW	apoptotic; osteopathic; antiarthritic; tranquiliser.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	/*tag= a
FT	/product= "PRSS11-L protein"
XX	
PD	US2004005659-A1.
XX	
XX	08-JAN-2004.
XX	
PF	03-JUL-2002; 2002US-00189099.
XX	
PR	03-JUL-2002; 2002US-00189099.
XX	

PA (DARR/) DARRON A. L.  
 PA (OLJ/) OL J.  
 PA (CHEN/) CHEN C.  
 PA (ANDR/) ANDRADE-GORDON P.  
 PI Darrow AL, OL J, Chen C, Andrade-Gordon P,  
 XX P-PSDB; ADJ11356.  
 DR WPI: 2004-081723/08.  
 DR P-PSDB; ADJ11356.  
 XX  
 PT New isolated S2 serine protease nucleic acid and polypeptides, useful  
 PT for diagnosing and/or treating diseases with aberrant expression or  
 PT activity the S2 serine protease, such as osteoarthritis, stress and  
 PT apoptotic disorders.  
 XX  
 PS Claim 2, SEQ ID NO 1; 28pp; English.  
 XX  
 CC This invention relates to a novel isolated nucleic acid molecule  
 CC identified as PRS11-L that encodes an S2 serine protease. Specifically,  
 CC it refers to members of the S2/Htra serine protease family, such that it  
 CC plays a role in cellular physiology and apoptosis. The present invention  
 CC provides agents, antagonists, antibodies and recombinant expression  
 CC vectors useful in methods of treatment, or detection and diagnosis of  
 CC diseases associated with the aberrant expression or activity of the S2  
 CC serine protease, PRS11-L. Accordingly, compositions described herein can  
 CC be used via gene therapy routes to treat osteoarthritis, stress and  
 CC apoptotic disorders. As such, they exhibit osteoprotective, antiarthritic and  
 CC tranquiliser activities. This polynucleotide sequence is the human PRS11  
 CC -L cDNA sequence of the invention.  
 CC  
 XX  
 SQ Sequence 3006 BP; 615 A; 918 G; 911 G; 562 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 9.03e-164 Length: 3006  
 Score: 1679.00 Matches: 334  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 12 Gaps: 0  
 US-10-617-443B-2 (1-334) x ADJ11355 (1-3006)  
 QY 1 Mech1aLeuAlaLeuProAlaSerAlaGlyLeuH1eGlnLeuSerSerProAlaGlyTys 20  
 DB 1011 ATGCACGTGGCCCTTCCCGCCAGCGACGATCTCCACAGCTGAGCGCCGCGCTCAAG 1070  
 QY 21 PheAspPhe11eAlaAspValValGlyVal1eAlaProAlaValValHis1eGlnLeu 40  
 DB 1071 TTCACCTTCATGCTGACGCTGCGAGAGATGCGACCGCTGCTCCACATAGAGCTC 1130  
 QY 41 PheLeuArgHisProLeuPheGlyArgAsnValProLeuSerSerSerGlyPhe11e 60  
 DB 1131 TTCCTAGACACCCCGCTGTTGGCCGACAGTCCCTGTCGACGCGTTCGCTTCATC 1190  
 QY 61 MetSerGluAlaGlyLeuLeu1e1eThrAsnAlaHisValValSerSerAsnSerAla1a 80  
 DB 1191 ATGTGACAGCGCCGCTGATCATCAACATGCCACGTGGTGTCCAGAACAGTGTGCC 1250  
 QY 81 ProGlyArgGlnGlnLeuValGlnLeuGlnAsnGlyAspSerTyrGluAlaThr11e 100  
 DB 1251 CCGGGAGCGACAGCACTCAAGTGCAGCTACAGAAATGGGGATCCCTATGAGGCCACATC 1310  
 QY 101 LysAspP1eAspPylsSerAspP1eAlaThr11eVal1eHisProValValValLeu 120  
 DB 1311 AAAGACATGACAGAAAGTGGACATTCACACATCAAGATCCCAAGAAAAGCTC 1370  
 QY 121 ProValLeuLeuLeuGlnHisSerAlaAspLeuArgProGlyGluPheValValAla1e 140  
 DB 1371 CCGGTGTGTGTGGGTCACTCGCGCGACCTGCGGCGCTGGGGAGTTTGTGGTGCATC 1430  
 QY 141 GlySerProPheAlaLeuGlnAsnThrValThrThGly11eValSerThrAlaGlnArg 160  
 DB 1431 GGCAAGTCCCTTCCCTACAGAACACAGTGCACAAAGCGGATGTCAGCATCGCCACGGG 1490

QY 161 GlnGlyArgGlnLeuGlnLeuArgAspSerAspMetAspTyr11eGlnThrAspAla1e 180  
 DB 1491 GAGGAGAGGAGACTGAGCTCGGAGACTCCGACATGACATCAACAGGATCCATC 1550  
 QY 181 IleAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGlnVal11eGly11e 200  
 DB 1551 ATCAACTACGGGAACTCCGGGGGACCACTGGTGAACCTTGATGGAGAGTATTGGCATC 1610  
 QY 201 AsnThrLeuValValThrAlaGly11eSerPheAla11eProSerAspArg11eThrArg 220  
 DB 1611 AACACGCTCAAGTACCGCTGGCATCTCTTGGCCATCCCTCAGACCGCATCACAGG 1670  
 QY 221 PheLeuThrGlnPheGlnAspPylsGln11eLysAspTyrPylsValArgPhe11eGly11e 240  
 DB 1671 TTCCTCACAGACTTCAGAACACAGATCAAGACTGAGAACGCGCTTCATCCGCATTA 1730  
 QY 241 ArgMetArgThr11eThrProSerLeuValAspGlnLeuLysAlaSerAsnProAspPhe 260  
 DB 1731 CGGATCGGAGACGATCAACCAAGCTGTGGATGAGTGAAGCGCAGCAACCGGACTTC 1790  
 QY 261 ProGluValSerSerGlyTyrValGlnGlnValAlaProAsnSerProSerGlnArg 280  
 DB 1791 CCAGAGGTGACGACGATGGAATTTATGTGCAGAGGTTGGCCGGAATTCACCTTTCAGAGA 1850  
 QY 281 GlyGly11eGlnAspGlyAspP1e11eValLysValAsnGlyArgProLeuValAspSer 300  
 DB 1851 GCGGACATCCAGATGAGTGAATCATCATCTCAAGATCAAGCGGCTCTTATGTGACTCG 1910  
 QY 301 SerGlnLeuGlnGlnAlaValLeuThrGlnSerProLeuLeuGlnValArgArg11y 320  
 DB 1911 AGTGAAGTGCAGAGAGCGGTGCTGACCGAGTCTCTCTCTTCACTGAGAGTCCGCGGGGG 1970  
 QY 321 AsnAspAspLeuLeuPheSer11eAlaProGluValValMet 334  
 DB 1971 AACGACGACCTTCTTTCAGCATGCGACCTGAGGTGTCATG 2012  
 RESULT 2  
 AAD44189  
 ID AAD44189 standard; cDNA, 1044 BP.  
 XX  
 AC AAD44189;  
 XX  
 DT 13-DEC-2002 (first entry)  
 XX  
 DE Mouse TANGO 219 cDNA.  
 XX  
 KW TANGO protein; vaccine; chromosomal mapping; antisense gene therapy;  
 KW forensic biology; predictive medicine; cytostatic; pharmacogenomic;  
 KW tissue typing; mouse; gene; ss.  
 XX  
 OS Mus musculus.  
 XX  
 PN US2002055139-A1.  
 PD 09-MAY-2002.  
 XX  
 PF 01-MAR-2001; 2001US-00796858.  
 XX  
 PR 30-DEC-1998; 98US-00223094.  
 PR 30-DEC-1998; 98US-00223546.  
 PR 30-DEC-1998; 98US-00224246.  
 PR 14-MAY-1999; 98US-00312359.  
 PR 18-JUN-1999; 98US-00336536.  
 PR 29-JUN-1999; 98US-00342687.  
 PR 30-JUL-1999; 98US-00365164.  
 PR 20-SEP-1999; 98US-00399723.  
 PR 23-DEC-1999; 98US-00471179.  
 PR 29-DEC-1999; 98US-00474071.  
 PR 29-DEC-1999; 98US-00474072.  
 PR 15-MAY-2000; 2000US-00572002.  
 PR 19-JUN-2000; 2000US-00597993.  
 PR 22-JUN-2000; 2000US-00599596.

PR 29-JUN-2000; 2000US-00606565.  
 PR 31-JUL-2000; 2000US-00630334.  
 PR 20-SEP-2000; 2000US-00665666.  
 XX  
 PA (HOLT/) HOLTZMAN D A.  
 PA (SHAR/) SHARP J D.  
 PA (LEIB/) LEIBY K R.  
 PA (BOSS/) BOSSONE S.  
 PA (PANY/) PAN Y.  
 PA (BARN/) BARNES T M.  
 PA (FRAS/) FRASER C C.  
 PA (WRIE/) WRIGHTON N.  
 PA (MYER/) MYERS P S.  
 PA (KING/) KINGSBURY G.  
 XX  
 PI Holtzman DA, Sharp JD, Leiby KR, Bossone S, Pan Y, Barnes TM;  
 PI Fraser CC, Wrighton N, Myers PS, Kingsbury G;  
 XX  
 DR WPI; 2002-453953/48.  
 XX  
 PT New isolated nucleic acid encoding a TANGO polypeptide, for use as a  
 PT modulating agent for regulating cellular processes and for use in a  
 PT vaccine.  
 XX  
 PS Claim 1; Page 81-82; 206pp; English.  
 XX  
 CC The invention relates to TANGO polypeptide and its corresponding nucleic  
 CC acid sequence. TANGO protein is used to identify a compound which binds  
 CC to it that can be used for modulating its activity. It is also used to  
 CC produce an antibody. The antibody is used to detect the presence of the  
 CC polypeptide in a sample. TANGO DNA and protein are useful as modulating  
 CC agents in regulating cellular processes. They can be used in vaccines.  
 CC TANGO DNA and protein and its antibody are used in e.g. chromosomal  
 CC mapping, tissue typing, forensic biology, predictive medicine,  
 CC pharmacogenomics and treatment methods. TANGO DNA is used in antisense  
 CC gene therapy. The present sequence is mouse TANGO 219 cDNA. Note: This  
 CC sequence SEQ.ID.NO:9 is said to encode the protein AAE26441, but this is  
 CC not the case  
 XX  
 XX Sequence 1044 BP; 230 A; 318 C; 306 G; 190 T; 0 U; 0 Other;  
 SQ  
 Alignment Scores:  
 Pred. No: 7,83e-160 Length: 1044  
 Score: 1635.00 Matches: 325  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 97.38% Indels: 0  
 DB: 6 Gaps: 0  
 US-10-617-443B-2 (1-334) x AAD44189 (1-1044)  
 QY 10 GlyLeuHisGlnLeuSerSerProArgTyrIlePheAsnPhelIeAlaAspValIaGlu 29  
 DB 70 GGTCTCCACGAGCTGACAGCCCGCTACAACTTCATTCATTCGTCGAG 129  
 QY 30 LysIleAlaProAlaValIaHisIleGlnLeuPheLysPheProLeuPheGlyArg 49  
 DB 130 AAGATCCACAGCGGTGTCACATAGAGTCTTCTCGAACAACCCGCTGTGGCCG 189  
 QY 50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGlnAlaGlyLeuIleIleThr 69  
 DB 190 AACGTGCCCTGTCTCCACCGGTCTGCTCATCTGTCAGAGCCGCGCTGATATACCC 249  
 QY 70 AsnAlaHisValIaSerSerAsnSerAlaAlaProGlyArgGlnIleuValGln 89  
 DB 250 AATGCCACCGTGTGTCCAGCAACAGTGTGCCCGGCGAGGCGAGCTCAAGTGCAG 309  
 QY 90 LeuGlnAsnGlyAspSerTyrGlnAlaThrIleLysAspIleAspIleValSerAspIle 109  
 DB 310 CTATGATAGAGGAGCTCTATGAGGCCACATCAAAACATCCGCAAGAGTCCGACATT 369  
 QY 110 AlaThrIleLysIleHisProLysIleValLeuProValIleuLeuAlaGlyHisSerAla 129

DB 370 GCCACCATCAAGATCCATCCCAAGAAAAGCTCCCTGTGTGTGTGTCGTGACTCGGCC 429  
 QY AspLeuArgProGlyIleValIleAlaIleGlySerProPheAlaLeuGlnAsnThr 149  
 DB 430 GACCTGGCGCTGGAGAGTTGTGTGCGCATTCGCGAGTCCCTTCGCTTACAGAACCA 489  
 QY 150 ValThrThrGlyIleValSerThrAlaGlnArgIleGlyLeuGlyLeuArgAsp 169  
 DB 490 GTGACAAAGGAGCATCTGTCAGACATCCAGCCGAGGAGGAGGAGTCCGCTCCGGAC 549  
 QY 170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyPro 189  
 DB 550 TCCGACATGAGCTACATCCAGACCGAGTGCATCATCACTCAAGGAACTCCGGGGAGCA 609  
 QY 190 LeuValLeuLeuAspGlyIleValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209  
 DB 610 CTGTGTGAACCTGGATGCGAGGTGCTATTGGCATCAACGCTCAAGGTCAACGCTGGCATC 669  
 QY 210 SerPheAlaIlePheSerAspArgIleThrArgPheLeuThrGlnAspIleGln 229  
 DB 670 TCCCTTGCCATCCCTCAGACCGCATCAGACGTTCTCTCAGAGTTCCAAAGACAGCAG 729  
 QY 230 IleLysAspTyrIleValArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249  
 DB 730 ATCAAGACTGGAAGAGCGCTTCATCGCATACGATGCGAGATCAACCAAGCCTG 789  
 QY 250 ValAspIleuLeuValAsnSerAsnProAspPheProGlyValSerSerGlyIleTyrVal 269  
 DB 790 GTGATAGCTGAAGGCGACACACCCGAGCTTCCCAAGGTGACAGATGAAATTATGTG 849  
 QY 270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIle 289  
 DB 850 CAAAGGTTCCGCGAATTACCTTCTCAGAGAGCGCGCATCCAAAGTGTGATCATC 909  
 QY 290 ValLysValAsnGlyArgProLeuValAspSerSerGlyLeuGlnGlnAlaValIleThr 309  
 DB 910 GTCAAGGTCAACGGGCTCTCTAGTGACTCGAGTGAAGTGCAGAGGCGGTGCTGACC 969  
 QY 310 GluSerProLeuLeuGlnValArgArgGlyAsnAspLeuLeuPheSerIleAla 329  
 DB 970 GAGTCTCTCTCTCTACAGAGGTGCGCGGGGGAACGACACCTCTCTTCAGATCGCA 1029  
 QY 330 ProGluValIleMet 334  
 DB 1030 CCTGAGGTGCTCATG 1044  
 RESULT 3  
 ADO01101  
 ID ADO01101 standard; DNA; 1316 BP.  
 AC ADO01101;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE Human serine protease HTRA3 related DNA, SEQ ID 16.  
 XX  
 KW Cytostatic; Serine protease inhibitor; Apoptosis stimulator; human;  
 XX serine protease; enzyme; HTRA3; pancreatic cancer; cancer; ds.  
 OS Homo sapiens.  
 XX  
 PN NC02064039407-A1.  
 XX  
 PD 13-MAY-2004.  
 XX  
 PF 30-OCT-2003; 2003MO-JP013920.  
 XX  
 PR 01-NOV-2002; 2002JP-00320075.  
 PR 27-JAN-2003; 2003JP-00017892.  
 XX  
 PA (TAKE) TAKEDA CHEM IND LTD.  
 XX  
 PI Horikoshi K, Kitahara O, Watanabe T, Taniyama Y, Nishizawa S;

XX WPI, 2004-400073/37.

XX Apoptosis inducers inhibiting activity or expression of serine protease  
XX HTRA3 for treatment, prevention and diagnosis of pancreatic cancer.

XX Example 3, SEQ ID NO 16, 124bp; Japanese.

XX The present invention relates to apoptosis inducers and protease  
XX inhibitors, which contain a compound inhibiting the activity or  
XX expression of human serine protease HTRA3 (AD001090 and AD001115). The  
XX invention is useful for the treatment, prevention and diagnosis of  
XX pancreatic cancer, and also of other cancers including colon, breast,  
XX lung, prostate, esophagus, stomach, liver, spleen, kidney, gall bladder,  
XX ovary, testis, bladder, thyroid and brain cancer and blood neoplasms. The  
XX present sequence was used to illustrate the invention.

XX Sequence 1316 BP; 254 A; 425 C; 404 G; 233 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	Length:	1316
Score:	1635.00	Matches: 325
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	97.38%	Indels: 0
DB:	12	Gaps: 0

US-10-617-443B-2 (1-334) x AD001101 (1-1316)

```

QY 10 GYLEHNGINLSESRERPROARGTYRLYSPHEASPHETLEAALAPVALVALGLU 29
DB 334 GGTCTCCACGAGTGAGACCCGCGCTACAGATTCACTTCTTGCTGACGTGGTGAG 393
QY 30 LYSILEALAPROALAVAlVALHISLEGLULPHLEUARGHISERPROLEUPHEG1YARG 49
DB 394 AATATGCGACCAACCCGCGTCCATATAGAGCTCTTCTTGAGACACCCGCTGTGGCCGC 453
QY 50 AENVALPROLEUSERSERGLYSERGLYPHEILEMERSERGLUALAGLYLEULEILETHR 69
DB 454 AACGTGCCCCCTGTCCAGCGGCTTCTGCTTCATATGCAGAGCCGCGCTGATCATCAC 513
QY 70 AANAHLHIVAlVALSERSEASERAlA1APROGLYARGINGINLEULYVALGLN 89
DB 514 AATGCCCACTGGTGTCCAGCAACAGTGTGCCCGGCGAGGACAGCTCAAGTGGCAG 573
QY 90 LEUGLNASNGLYASPSERYRGUUALATHRILEYASAPLLEAPLYSISERAPLLE 109
DB 574 CTACAGATGGGAGCTCTATGAGGCCACATCAAGACATCGAAGAAGTCGACATT 633
QY 110 ALATHRILEYALIEHSPOLYBLYSLEUPROVALLEUULEUENGLYHISERAlA 129
DB 634 GCCACCATCAAGATCCATCCAGAAAAAGCTCCGTGTGTGTGGTGGGTCACTCGGCC 693
QY 130 ASPLLEUARGPROGLYGLUPHEVALVALA1A1EGLYSERPROPHALALEUGINASNTHR 149
DB 694 GACCTGGGCGCTGGGAGTTTGTGTGGTGGCCATCGCGAGTCCCTTCCGCTACAGAAC 753
QY 150 VALTHRTIRGLYILEVALISERTHRA1AGLNAIRGLUIGLYARGULLEUGLYLEUARGAP 169
DB 754 GTGACACCGGAGCTCTGACACAGCCAGCGGAGGAGGAGCGGCGCTCCGGGAC 813
QY 170 SERAPMECLAPRYTIRLEGINTHRAAPALALIELEHENTYRGYASANSERGLYPRO 189
DB 814 TCCGACATGGACATCAATCCAGACGAGATGCATCTCACTACAGGAACTCCGGGGAGCCA 873
QY 190 LEUVALASNTLEUAPRG1YGLUVALI1EGLY1LEANTHRIEULYVAL1THRALAGLY1LE 209
DB 874 CTGGTGAACCTGATGGCGAGGTATGGCATCAACAGCTCAAGGTCAAGGCTGGCATC 933
QY 210 SERPHEALALIEPROSERAPARGILETHRARGPHELEUTHRG1UPHEGLINAPLYSGIN 229
DB 934 TCCTTTGCATCCCTCAGACCGCATCAACGAGTTCTCTCAAGAGTTCCAAAGCAG 993

```

```

QY 230 ILEYASAPTRPLYBLYSARGPHEILEGLYILEARGMEARGTHRIETHRPROSERLEU 249
DB 994 ATCAAGAGCTGAGAGAGCGCTTCAITCGCATACGATGCGGACATCAACAGAGCTTG 1053
QY 250 VALAEPGLULEUVALISERASNPAPROAPHPROGLUVALISERSEGLY1LETYVAL 269
DB 1054 GTGATGAGCTGAGAGCGGACGAGACCCGAGCTTCCAGAGAGTCAAGAGATTATGTG 1113
QY 270 GINGUVALA1APROANSERPROSERGLNARGGLYGLY1LEGINASGLYASPLLE1LE 289
DB 1114 CAAAGGTTGGCCCGCAATTCATCTTCAAGAGGCGGCAATCAAGATGTGATCAATTC 1173
QY 290 VALLYSVALASNGLYARGPROLEUVALASPSERSEGLYLEUGINGUVALAVALLEUTHR 309
DB 1174 GTCAAGGTCAACGGGCGTCTTCACTGATGACATCGAGTGCAGAGGCCGTGTCAC 1233
QY 310 GLUSERPROLEULEUENGLUVALARGARGGLYASNAAPLEULEUPHESERILEALA 329
DB 1234 GAGTCTCCTCTCTCACTGAGAGGTGCGGCGGGAACGACAGACCTCTTCAAGATCGCA 1293
QY 330 PROGLUVALVALMET 334
DB 1294 CCTGAGTGTGTCATG 1308

```

#### RESULT 4

AD001102

ID AD001102 standard; DNA; 1319 BP.

AC AD001102;

DT 29-JUL-2004 (first entry)

DE Human serine protease HTRA3 related DNA, SEQ ID 17.

KW Cytostatic; Serine protease inhibitor; Apoptosis stimulator; human;

OS Homo sapiens.

PN WO2004039407-A1.

PD 13-MAY-2004.

PF 30-OCT-2003; 2003WO-JP013920.

PR 01-NOV-2002; 2002JP-00320075.

PR 27-JAN-2003; 2003JP-00017892.

PA (TAKE ) TAKEDA CHEM IND LTD.

PI Horikoshi K, Kitahara O, Watanabe T, Taniyama Y, Nishizawa S;

DR WPI, 2004-400073/37.

PT Apoptosis inducers inhibiting activity or expression of serine protease

PT HTRA3 for treatment, prevention and diagnosis of pancreatic cancer.

XX Example 3, SEQ ID NO 17, 124bp; Japanese.

XX The present invention relates to apoptosis inducers and protease  
XX inhibitors, which contain a compound inhibiting the activity or  
XX expression of human serine protease HTRA3 (AD001090 and AD001115). The  
XX invention is useful for the treatment, prevention and diagnosis of  
XX pancreatic cancer, and also of other cancers including colon, breast,  
XX lung, prostate, esophagus, stomach, liver, spleen, kidney, gall bladder,  
XX ovary, testis, bladder, thyroid and brain cancer and blood neoplasms. The  
XX present sequence was used to illustrate the invention.

XX Sequence 1319 BP; 255 A; 425 C; 405 G; 234 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	Length:	1319
Score:	1635.00	Matches: 325

Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.38%	Indels:	0
DB:	12	Gaps:	0

US-10-617-443B-2 (1-334) X ADO01102 (1-1319)

10 Glycylglycyl-L-leucine

```

Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

```

-1319)

ArgTyrLysPheAsnPh

CGCTACCAAGTTCACCTT

Qy		30	LysIlealProalaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg	49
Db		394	AAGATCGACCAAGCCGTGGTCCAATTAGAGCTTCCTTAGACACACCCTGTTGGCCGC	453
Qy		50	AsnValProLeuSerSerSeriYseSrglyPheIIemebSerGluAlaGlyLeuIleIethr	69
Db		454	AACGGCCCCCTGTCCAGCGGATTCTGGCTTCATCATGTCAAGAGGCCGCTGATCATCAC	513
Qy		70	AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGluLeuYsValGln	89
Db		514	AATGCCCAAGTGAGTGTCCAGCAACAAGTGTGCCCGGGAGGAGCAGAGCTCAAGGTGCA	573
Qy		90	LeuGlnaEngLYAaspSerTyrgLyuAlaThrIleYsAspIleAAspyYsSeRaSpiIe	109
Db		574	CTACAGAATGGGAATCTCTATAGGCCAACCATCAAGATGACAAAGAGTCCGACATT	633
Qy		110	AlaThrIleYsIleHisProYsYsYsLeuProValLeuLeuLeuGlnYhisSerAla	129
Db		634	GCCACCATCAAGATCCATCCCAAGAAAAGCTCCCTGTGTGTGTGGTACTCGGCC	693
Qy		130	AspLeuArgProGlyYglUnpHeValValAlaleGlySerProPheAlaLeuGlnaAnthr	149
Db		694	GACCTGCGCCCTGGGAGATTGTGTGGTCATCGGACAGTCCCTTCCTACAGAACAA	753
Qy		150	ValThrThrArgLyIleValSerThrAlaGlnArgGluArgGluLeuEngLYeuaArgAp	169
Db		754	GTCACACAGGGCATGTCTACACCTGCCAGCGGAGGCGAGGAGCTGGAGCTCCGGAGC	813
Qy		170	SerAspMetAspTYrIleGlnThrAspAlaIleIleAsnTYrGlyYAasnserGlygPro	189
Db		814	TCCGACATGACCTACATCCAGCGGATGCCATCATCACTAAGGAACTCCGGGGAGCA	873
Qy		190	LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuYsValThrAlaGlyIle	209
Db		874	CTGGTGAACCTCGAAGGCGAGGTCAATTGGCATCAACACCTCAAGGTCAAGCTGGCATC	933
Qy		210	SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspYsGln	229
Db		934	TCCTTTGCATATCCCTCAGACCGCATCAACGCGTTCCTACAGAGTCCMAAGCAAGCAG	993
Qy		230	IleYsAspTYrPLySlyAspPheIleGlyIleArgMetArgThrIleThrProSerLeu	249
Db		994	ATCAAGAAGCTGGAGAAAGCGCTTCAATCCGATACGGATGCCAGCATCAACCAAGCTCG	105
Qy		250	ValAspGluLeuYsAlaSerAsnProAspPheProGluValSerSerGlyIleTYrVal	269
Db		1054	GTGGATGAGCTGAAGGCCAGCAACCGGACTTCCCAGAGGTCAAGCTGAATTTATGTG	111
Qy		270	GlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnAspGlyAspIleIle	289
Db		1114	CMAAGGTGGCGCCCAATTCACCTTCTCAGAGAGCGGCGATCAAGATGGTGAATCATCATC	117
Qy		290	ValYsValaAngLYArgProLeuValAspSerSerGluLeuGlnGluValaLeuThr	309
Db		1174	GTCAGAGTCAACGGCGCTCTCTTAATGAGACTCAAGTGAAGCTCAAGAGGCCGTGCTGAC	123
Qy		310	GluSerProLeuLeuLeuGluValaArgArgGlyAsnAspAspLeuLeuPheSerIleAla	329
Db		1234	GAGTCTCTCTCTTAATGAGGTGGCGGGGGAACAGACCTCTCTTCAAGCATCGCA	129
Qy		330	ProGluValValMet	334

QY	50	AsnValaProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr	69
Db	505	AAAGTGCCCTGTCGACAGGGGTTCTGGCTTCATATGTCACAGGCGGCTGATCATCACCC	564
QY	70	AsnAlaHisValSerSerAsnSerAlaAlaProGlyArgGlnLeuIleValGln	89
Db	565	AAATGCCACCTGGTGTCACCAACAGATGCTGCCCGGGCAGGACGCTCAAGGTGCGAG	624
QY	90	LeuGlnAsnGlyAspSerTyrGlnAlaIleThrIleLeuAspIleLeuPheValysSerAspIle	109
QY	625	CTACAGAAATGGGGCACTCTATGAGGCCACCATCAAAAGCATCGACAAGATGGACATT	684
Db	110	AlaThrIleLeuSerIleHisProLeuLeuLeuProValLeuLeuLeuGlyHisSerAla	129
Db	685	GCCACCATCAAGATTCATCCCAAGAAAAGCTCCCTGTGTGTTGTTGCTGGGCTCACTCGGCC	744
QY	130	AspLeuAsnArgProGlyGluPheValAlaAlaIleGlySerProPheAlaIleuGlnAsnThr	149
Db	745	GACCTGGGGCTGGGGAGTTGTGTGTGGCCATCGGAGTCCCTTGCTGCTGGGCTCACTCGGCC	804
QY	150	ValThrThrArgIleIleValSerThrAlaGlnArgGlnGlyArgGlnLeuGlyLeuAsnAsp	169
Db	805	GTGACAAAGGAGCATCTGTACACACTGCCAGGGGAGGACAGGAGGCTGGGGCTCGGGAGC	864
QY	170	SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro	189
Db	865	TCCGACATGAGCATCATCATCCAGACGGATGCCCATCATCACTACGGGAACTCCGGGGAGCCA	924
QY	190	LeuValAsnLeuAspGlyGlyValIleGlyIleAsnThrLeuIleValThrAlaGlyIle	209
Db	925	CTGTGAAACCTGGATGGGAGGTCAATTGGCATCAACAGCTCAAGGTCAAGCTGGCATC	984
QY	210	SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGlnPheGlnAspValysGln	229
Db	985	TCTTTGGCAATCCCTCTAGACCGCATCACGGTTCCTCACAGAGTCCAAAGACAAGCAG	1044
QY	230	IleLeuAspTyrPheLeuArgPheIleGlyIleArgMetArgThrIleThrProSerLeu	249
Db	1045	ATCAAGACTGGAAAGACGGCTTCATCCGCATACGGATGCCAGATCAGATCACAAGCCTG	1104
QY	250	ValAspGluLeuLeuAlaSerAsnProAspPheProGluValSerSerGlyIleThrVal	269
Db	1105	GTGATGTAGCTGAGMAAGCCGACCAACCGGACTTCCAGAGGTGACGATGGAAATTATATG	1164
QY	270	GlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnAspGlyAspIleIle	289
Db	1165	CAGAAGGTGGCCCGCAATTCACCTTCTCGAAGAGGGGCATCAAGATGTGTACATCATC	1224
QY	290	ValIleValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThr	309
Db	1225	GTCAAGAGTCAAGGGGGTCTCTCTAGTGCATCTGAGTGAAGCTGACAGAGGCGCTGGTCAAC	1284
QY	310	GluSerProLeuLeuLeuGlnValAlaArgArgGlyAsnAspAspLeuLeuPheSerIleAla	329
Db	1285	GAGTCTCTCTCTCTACTGAGGTGGCGGGGGGAAGAGACCTCTCTTGACATCGCA	1344
QY	330	ProGluValValMet	334
Db	1345	CCTGAGGTGTGTCATG	1359
RESULT 6			
AD001099			
ID AD001099 standard; DNA; 1375 BP.			
XX			
AC AD001099;			
XX			
DT 29-JUL-2004 (first entry)			
XX			
DE Human serine protease HTRA3 related DNA, SEQ ID 14.			
XX			
CY Cysteatic; Serine protease inhibitor; Apoptosis stimulator; human;			
KW serine protease; enzyme; HTRA3; pancreatic cancer; cancer; de.			

XX	Homo sapiens.
XX	WO2004039407-A1.
PN	
XX	13-MAY-2004.
PD	
XX	30-OCT-2003; 2003WO-JP013920.
XX	
PF	01-NOV-2002; 2002JP-00320075.
XX	
PR	27-JAN-2003; 2003JP-00017892.
XX	
PA	(TAKE ) TAKEDA CHEM IND LTD.
XX	
PI	Horikoshi K, Kitahara O, Watanabe T, Taniyama Y, Nishizawa S;
XX	
DR	WPI, 2004-400073/37.
XX	
PT	Apoptosis inducers inhibiting activity or expression of serine protease
XX	HTK33 for treatment, prevention and diagnosis of pancreatic cancer.
PS	
XX	Example 3; SEQ ID NO 14; 124bp; Japanese.
CC	The present invention relates to apoptosis inducers and protease
CC	inhibitors, which contain a compound inhibiting the activity or
CC	expression of human serine protease HTK33 (AD001099 and AD001115). The
CC	invention is useful for the treatment, prevention and diagnosis of
CC	pancreatic cancer, and also of other cancers including colon, breast,
CC	cervix, prostate, esophagus, stomach, liver, spleen, kidney, gall bladder,
CC	ovary, testis, bladder, thyroid and brain cancer and blood neoplasms. The
CC	present sequence was used to illustrate the invention.
XX	
SO	Sequence 1375 BP; 258 A; 447 C; 429 G; 241 T; 0 U; 0 Other;
Alignment Scores:	
Pred. No. :	1,14e-159      Length:      1375
Score:	1635.00      Matches:      325
Percent Similarity:	100.00%      Conservative:      0
Best Local Similarity:	100.00%      Mismatches:      0
Query Match:	97.38%      Indels:      0
DB:	12      Gaps:      0
US-10-617-443B-2 (1-334) x AD001099 (1-1375)	
OY	10 G1YLEUHIAGLNFLEUSESEPROARYTYLYSPHEASNPHEILALANPVAlValGlu 29
Dd	393 GGCTTCACACAGGTGACGACC GCCGCTCAAGTTCAATTCATTCTCAACGGTGGAG 452
OY	30 LysILEAIProlAVAlValHISIIeGLULEUPHeLAURghIAPROleUPheGIYArg 49
Dd	453 AAATGGCACCAACCGTGTCACATAGACTTTCTTAGAKAACCCGGCTTTGGCCGC 512
OY	50 AsnVALProLeUSESeRGISerGIYPHeIIemETSerGIUALagLYLeuIIelIEthr 69
Dd	513 AACGTGCCCTGTCCACGCGGTTCTGGCTTCATCATGTCAAGAGGCCGCGCTGATCATCACC 572
OY	70 AsnALAHisVALserSERanSeraIALAProGIYARgInGInLeuLVAlGln 89
Dd	573 AATGGCCACCGTGTCACAGCAACAGTGCCGCCGGACGGACNAGACTCAAGTCACAG 632
OY	90 LeuGINAsngIYAASPseRTYGIUALATHrIIeLYsApIIeasPYsLSerapIIe 108
Dd	633 CTACAGAAATGGGACTCTCTATGAAGGCCACATCAAAGACATGCAGAAGTCGGACATT 692
OY	110 ALATHrIIeLYsIIeHspOLYsLYsLYLeuLPVOAlleuLeuLengIYHISecALA 122
Dd	693 GCCACCACTCAAGATCCATCCCAAGAAAACTCCCTGTGTGTGTGTGGGTCACTCCGCC 75
OY	130 ASPLeuArRPProGIYGIUpHeVAIVAlAIleGLYSerPROPhEalALEuGINAsnthr 148
Dd	753 GAACCTGGGGCTGGGGAGTTTGTGTGTGTGGCATGGGAGTCCCTTCGCTTACAGAACACA 81
OY	150 ValThrThrgLYlIeValISerTHraIGlnARgInuclYARGIuLeuNGIyleuARgAsp 166

```

Db      813 GTGCAACGGGCATGTCAGACATGCCAGCGGGAGGCGAGCTGGGCTCCGGGAC 872
Qy      170 SerpentineAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyPro 189
Db      873 TCCGACATGAGCTACATCCACACGAGTCCATCATCACTACGGGAACTCCGGGGGACCA 932
Qy      190 LeuValAsnLeuAspGlyValIleGlyIleAsnThrLeuValThrAlaGlyIle 209
Db      933 CTGGTGAACCTGGATGGGAGGCTCATTTGGCATCAACGCTCAAGGTACCGGCTGGCATC 992
Qy      210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspGln 229
Db      993 TCCTTGGCATCCCTCAGACCGCATCACAGGTTCTCCACAGAGTTCCAAAGACAGAG 1052
Qy      230 IleValAspTyrPheValArgPheIleGlyIleArgMetAlaGlnThrIleThrProSerLeu 249
Db      1053 ATCAAAAGCTGGAAGAAGCGCTTCATCGCATACGAGATCGGAGCATCAACAGAGCTG 1112
Qy      250 ValAspGluLeuValAsaSerAsnProAspPheProGluValIleSerSerGlyIleTyrVal 269
Db      1113 GTGATGAGCTGAAGGCCAGCAACCGGACTTCCAGAGGTCAAGAGTGAATTTATGTG 1172
Qy      270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleIle 289
Db      1173 CAAGAGGTGGCCGAATTCACTTTCAGAGAGCGGCATCCAAGATGTGATCATC 1232
Qy      290 ValIleValAsnGlyValArgProLeuValAspSerSerGluLeuGlnGluValLeuThr 309
Db      1233 GTCAAGGTCAACGGGCGCTCTCTAGTGACTCGAGTGGAGTGGAGGCGGCTGTGCC 1292
Qy      310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329
Db      1293 GAGTCTCTCTCTCTACATGGAGAGTGGGGGGAACGACGACTCTCTTCAAGATGCA 1352
Qy      330 ProGluValValMet 334
Db      1353 CTGAGGTGGTCATG 1367

RESULT 7
AD001093
ID      AD001093 standard; DNA; 1378 BP.
XX
AC      AD001093;
XX
DT      29-JUL-2004 (first entry)
XX
DE      Human serine protease HTRA3 related DNA, SEQ ID 8.
XX
KW      Cytosolic; Serine protease inhibitor; Apoptosis stimulator; human;
XX      Serine protease; enzyme; HTRA3; pancreatic cancer; cancer; ds.
XX
OS      Homo sapiens.
XX
PN      WO2004039407-A1.
XX
PD      13-MAY-2004.
XX
PF      30-OCT-2003; 2003WO-JP013920.
XX
PR      01-NOV-2002; 2002JP-00320075.
XX
PR      27-JAN-2003; 2003JP-00017892.
XX
PA      (TAKE ) TAKEDA CHEM IND LTD.
XX
PI      Horikoshi K, Kitahara O, Watanabe T, Taniyama Y, Nishizawa S;
XX      WPI; 2004-400073/37.
XX
PT      Apoptosis inducers inhibiting activity or expression of serine protease
XX      HTRA3 for treatment, prevention and diagnosis of pancreatic cancer.
XX
PS      Example 3; SEQ ID NO 8; 124bp; Japanese.

```

```

XX      CC The present invention relates to apoptosis inducers and protease
XX      CC inhibitors, which contain a compound inhibiting the activity of
XX      CC expression of human serine protease HTRA3 (AD001093 and AD001115). The
XX      CC invention is useful for the treatment, prevention and diagnosis of
XX      CC pancreatic cancer, and also of other cancers including colon, breast,
XX      CC lung, prostate, oesophagus, stomach, liver, spleen, kidney, gall bladder,
XX      CC ovary, testis, bladder, thyroid and brain cancer and blood neoplasms. The
XX      CC present sequence was used to illustrate the invention.
XX
SQ      Sequence 1378 BP; 259 A; 447 C; 430 G; 242 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 1,156-159 Length: 1378
Score: 1635.00 Matches: 325
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.38% Indels: 0
DB: 12 Gaps: 0
XX
US-10-617-443b-2 (1-334) x AD001093 (1-1378)
Qy      10 GlyLeuHisGlnLeuSerSerProArgTyrLeuPheAsnPheIleAlaAspValValGlu 29
Db      393 GGTCTCCAGCAGCTGAGACAGCCCGGCTTCAAGTTCACTTCAATGCTGACGTGTGGAG 452
Qy      30 LysIleAlaProAlaValIleHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
Db      453 AAGATCCACACAGCGCGGTGCCATGAGACTTCTCTGAGACACCGCTGTGGCGCG 512
Qy      50 AsnValProLeuSerSerSerGlySerGlyPheIleMetSerGluIleGlyLeuIleThr 69
Db      513 AACGTGCCCTGTCTCCAGCGGTCTGTGCTTCAATGTCAGAGCGGCGCTGATATACCC 572
Qy      70 AsnAlaHisValIleSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuValGln 89
Db      573 AATGCCACAGTGGGTGCACCAACAGTGTGCCCGGAGGAGGACGACGTCAAGTGTGAC 632
Qy      90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLeuAspIleAspIleValSerAspIle 109
Db      633 CTACAGAAATGGGAACTCTCTATGAGGCGCACCATCAAGACATCAAGAAAGTCGACATT 692
Qy      110 AlaThrIleLeuValIleHisProLeuValLeuProValLeuLeuLeuGlyHisSerAla 129
Db      693 GCCACCATCAAGATCCATCCCAAGAAAGCTCCCTGTGTGTGTGGTCACTCGGCC 752
Qy      130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
Db      753 GACCTGGCGGCTGGGAGTTTGTGTGGCATGGGAGTCCCTTCCGCTTACAGAACACA 812
Qy      150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyValArgGluLeuValArgAsp 169
Db      813 GTGCAACGGGCATGTCAGACATGCCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 872
Qy      170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyPro 189
Db      873 TCCGACATGAGCTACATCCACACGAGTCCATCATCACTACGGGAACTCCGGGGGACCA 932
Qy      190 LeuValAsnLeuAspGlyValIleGlyIleAsnThrLeuValThrAlaGlyIle 209
Db      933 CTGGTGAACCTGGATGGGAGGCTCATTTGGCATCAACGCTCAAGGTACCGGCTGGCATC 992
Qy      210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspGln 229
Db      993 TCCTTGGCATCCCTCAGACCGCATCACAGGTTCTCCACAGAGTTCCAAAGACAGAG 1052
Qy      230 IleValAspTyrPheValArgPheIleGlyIleArgMetAlaGlnThrIleThrProSerLeu 249
Db      1053 ATCAAAAGCTGGAAGAAGCGCTTCATCGCATACGAGATCGGAGCATCAACAGAGCTG 1112
Qy      250 ValAspGluLeuValAsaSerAsnProAspPheProGluValIleSerSerGlyIleTyrVal 269
Db      1113 GTGATGAGCTGAAGGCCAGCAACCGGACTTCCAGAGGTCAAGATGGAATTTATGTG 1172

```



QY	270	GINGLUVALALAPROANSERPROSERGINARGLVGLIEGINAEPGLAASPILLE	289
DB	1173	CMAAGAGTGGCCGGAATTCCTTCTCAGAGAGCGGCAATCCAGATGGACATCATC	1232
QY	290	VALLYSVLAANGLYARGPROLEUVAlASPERSERGIULEUGINGUVALIAVALLIETHR	309
DB	1233	GTCAAGGTCAACGGCGCTCTCTAGTGAGCTCGACTGACGTGCAAGAGCGCGTGTGACC	1292
QY	310	GLUSEPRPROLEULEULEUGLUVAlARGARGGLYANASAPAPPLAULEUPHESERILEA	329
DB	1293	GAGTCTCCTCTCCCTCACTGAGAGGTGCGGGGGAACGACGACTCTCTTCAGCATCCCA	1353
QY	330	PROGLUVALIMET	334
DB	1353	CCTGAGGTGCTCATG	1367
RESULT 8			
ID	AAA57361	standard; cDNA; 1563 BP.	
XX	AAA57361;		
XX	03-OCT-2000	(first entry)	
DE	cDNA encoding a Htra-2 (high temperature requirement A-2) protein.		
XX	Htra-2;	high temperature requirement A-2 protein; TANGO 219; arthritis;	
KW	bone disorder;	cartilage disease; bone breakage; cartilage tearing;	
KW	osteoporosis;	osteoarthritis; bone wearing; ischemic heart disease;	
KW	atherosclerosis;	hyperplasia; ventricular fibrillation; visceromegaly;	
KW	placentomegaly;	cardiac defect; Beckwith-Wiedemann syndrome; cancer;	
KW	cellular migration;	cell proliferation; serine protease;	
KW	Alzheimer's disease;	multiple sclerosis; Huntington's disease;	
KW	Parinson's disease;	neurodegeneration; frontal lobe dementia;	
KW	cortical lewy body disease;	acquired immunodeficiency syndrome; AIDS;	
KW	demential; chronic epilepsy;	adipocyte function; heart disease;	
KW	ischemic heart disease;	ventricular fibrillation; myocardial infarction;	
OS	89.		
XX	Mus sp.		
XX	Location/Qualifiers		
PH	Key		
FT	CDS	268..1314	
FT	/tag= a	/product= "Htra-2 (high temperature requirement A-2) protein"	
FT	/tag= b	replace(396, C)	
FT	/note=	"the encoded amino acid is changed to Asp; this variation is present only when the other two variations are absent"	
FT	variation	replace(426, C)	
FT	/tag= c	/note= "the encoded amino acid is changed to Asp; this variation is present only when the other two variations are absent"	
FT	variation	replace(498, C)	
FT	/tag= d	/note= "the encoded amino acid is changed to Asp; this variation is present only when the other two variations are absent"	
FT	variation		
XX	W0200039149-A2.		
XX	06-JUL-2000.		
XX	29-DEC-1999;	99WO-US031157.	
XX	30-DEC-1998;	98US-00224246.	
XX	(MILL-)	MILLENNIUM PHARM INC.	

PI Barnes TM;  
 XX WPI; 2000-452366/39.  
 DR P-PSDB; AAY93963.  
 DR  
 XX  
 XX Novel human high temperature requirement-2 and TANGO 219 secreted  
 PT proteins and nucleic acids encoding them which are useful for treating  
 PT atherosclerosis, placentalomegaly, cancer, Alzheimer's disease and  
 PT diabetes.  
 XX  
 XX Claim 2; Fig 8A-C; 122pp; English.

Claim 2; Fig 8A-C; 122pp; English

The present sequence encodes a murine HtrA-2 (high temperature requirement A-2) protein. The specification also describes TANGO 219 polypeptides. HtrA-2 nucleic acid, proteins and its modulators can be used to treat bone and/or cartilage associated diseases or disorders, e.g. bone breakage, cartilage tearing, osteoporosis, arthritis, osteoarthritis, and bone wearing. HtrA-2 is also used to treat ischemic heart disease, atherosclerosis, hyperplasia, and ventricular fibrillation. HtrA-2 can also modulate IGF function and can be used to treat vitseomegaly, placentomegaly, cardiac and adrenal defects, Beckwith-Wiedemann syndrome, and cancer. It is useful to treat disorders involving abnormal cellular migration, proliferation. HtrA-2 acts as a serine protease and can be used to treat Alzheimer's disease. HtrA-2 can also be used to treat multiple sclerosis, Huntington's disease, Parkinson's disease, neurodegeneration, frontotemporal dementia, cortical Lewy body disease, acquired immunodeficiency syndrome (AIDS), dementia and chronic epilepsy. HtrA-2 nucleic acids, proteins and its modulators can modulate adipocyte function and adipocyte-related processes and disorders. TANGO 219 can be used to treat heart diseases or conditions, e.g., ischemic heart disease or atherosclerosis, or cerebrovascular accidents, or more particularly, treating or preventing conditions involving heart contraction, and the impulse generating nodes and cardiac muscle cells, e.g. ventricular fibrillation or myocardial infarction

Sequence 1563 BP; 337 A; 462 C; 458 G; 306 T; 0 U; 0 Other

Alignment Scores:	
Pred. No.:	1,36e-159
Score:	1635.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	97.38%
DG:	3
Gaps:	0
Mismatches:	0
Conservative:	35
Indels:	0
Length:	1563

US-10-617-443B-2 (1-334) X AAA57361 (1-1563)

Qy	10	GIyleuHieglInleuSerSerProargIryLyvPheAsnPheIleAlaAspValValGlu	29
Db	337	GGITCCACACAGCTGAGCAGCCGCGCTCAAGTTCAACTTCATTTCCTGACGCGTGAGAA	396
Qy	30	LyvIleAlaProAlaValValHisIleGluPheLeuArgHisProLeuPheGlyArg	49
Db	397	AAAGATGCACACAGCCCTGTGTCCACAAGAGACTTCTCTGAGACACCCGCTGTTGGCCGC	456
Qy	50	AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr	69
Db	457	AAAGTCCCTCGTCCACAGCGGTTCTGGCTTCATATGTACAGAGCCGCGCTGATCATCACC	516
Qy	70	AsnAlaHisValValSerSerAsnSerAlaAlaProGlyValGlnGluLeuValGln	89
Db	517	AAATGCCACATGGTGTTCACCAACAAGTGTGCCCGGCGAGCAGCAGCTCAAGGTGCAG	576
Qy	90	LeuGlnAsnGlyAspSerIryGluAlaThrIleLeuAspIleAspIleAspIleAspIle	109
Db	577	CTACAGATGGGGACTCTCTATGAGGCGCACCATCAAAAGCATCAAGAAAGTCGACATT	633
Qy	110	AlaThrIleLeuValIleHisProGlyLeuLeuProValLeuLeuLeuGlyHisSerAla	128
Db	637	GCACGCCATCAAGATCCATCCCAAGAAAGAACTCCCTGTGTGTGTGCTGGGTCACTCGGCC	696
Qy	130	AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr	149



Db 697 GACCTCGGCGCTGGGAGCTTTGTGTGGCCATCGGACGCTCCCTTCGCTCAAGAACCA 756  
 Qy 150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyLeuArgAsp 169  
 Db 757 GTGACACGGGCGATCGTACGACCTCCAGGGGGGAGGAGCTGGGCTCCGGGAC 816  
 Qy 170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189  
 Db 817 TCCGACATGACATCATCCACACGGATCCATCATCACTACGGGAATCCGGGGGACCA 876  
 Qy 190 LeuValAsnLeuAspGlyValIleGlyIleAsnThrLeuValThrAlaGlyIle 209  
 Db 877 CTGTGTGAACCTGGATGGGAGGATCATTTGGCATCAACGCTCAAGTCAAGCTGGCATC 936  
 Qy 210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrIlePheGlnAspGlyGln 229  
 Db 937 TCCCTTGGCATCCCTCAGACCGCATCAACGGTTCTCCACAGAGTTCCAAAGACACAGAG 996  
 Qy 230 IleLysAspTyrIleValArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249  
 Db 997 ATCAAGACTGGAGAGGCGCTTCATCGGCATACGGATCGGACGATCACCAAGCTG 1056  
 Qy 250 ValAspGluLeuValAspAsnProAspPheProGluValSerSerGlyIleTyrVal 269  
 Db 1057 GTGGATGAGCTGAAGCGCAGCAACCCGACTTCCAGAGGTCAAGCATGGAATTTATGTG 1116  
 Qy 270 GlnGluValAlaProAsnSerProSerGlnArgIleGlyIleGlnAspGlyAspIleIle 289  
 Db 1117 CAAAGAGTGGCGCCGAATTCCTTCTCAGAGAGCGGCATCCAAAGTGGAGCATATC 1176  
 Qy 290 ValIleValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValIleThr 309  
 Db 1177 GTCAAGTCAACGGGCGCTCTCTAGTGACTCGAGTACGTCGACGAGGCGCTCTGACC 1236  
 Qy 310 GluSerProLeuLeuLeuGluValAlaArgGlyValAsnAspLeuLeuPheSerIleAla 329  
 Db 1237 GAGTCTCTCTCCACTGAGAGTGGCGGGGGAACGACGCTCTTTCAGCATCGCA 1296  
 Qy 330 ProGluValValMet 334  
 Db 1297 CCTGAGGTGGTCATG 1311  
 RESULT 9  
 ACF16993  
 ID ACF16993 standard; cDNA; 1583 BP.  
 AC ACF16993;  
 DT 15-SEP-2003 (first entry)  
 XX Human NOV24a protein encoding. cDNA SEQ ID NO:109.  
 DE Human NOV24a protein encoding. cDNA SEQ ID NO:109.  
 XX Human; NOVX; G protein-coupled receptor; cytosolic; cardiovascular;  
 KM immunosuppressive; anti-HIV; antiaslathmic; antiarteriosclerotic; AIDS;  
 KM hypotensive; gene therapy; cardiomyopathy; atherosclerosis; hypertension;  
 KM congenital heart defect; aortic stenosis; atrial septal defect; neoplasm;  
 KM arteriovenicular canal defect; pulmonary stenosis; prostate cancer;  
 KM uterine cancer; graft versus host disease; multiple sclerosis; GPCR;  
 KM acquired immunodeficiency syndrome; Crohn's disease; bronchial asthma;  
 KM chromosome mapping; forensic identification; gene; ss.  
 XX Homo sapiens.  
 OS  
 XX  
 PN W0200290568-A2.  
 XX  
 PD 14-NOV-2002.  
 XX  
 PF 02-MAY-2002; 2002W0-US014341.  
 XX  
 PR 03-MAY-2001; 2001US-0288935P.  
 PR 07-MAY-2001; 2001US-0289087P.  
 PR 08-MAY-2001; 2001US-0289620P.  
 PR 08-MAY-2001; 2001US-0289621P.

PR 09-MAY-2001; 2001US-0289817P.  
 PR 09-MAY-2001; 2001US-0289818P.  
 PR 11-MAY-2001; 2001US-0290194P.  
 PR 14-MAY-2001; 2001US-0290753P.  
 PR 15-MAY-2001; 2001US-0291189P.  
 PR 16-MAY-2001; 2001US-0291243P.  
 PR 18-MAY-2001; 2001US-0292001P.  
 PR 21-MAY-2001; 2001US-0292374P.  
 PR 22-MAY-2001; 2001US-0292587P.  
 PR 23-MAY-2001; 2001US-0293107P.  
 PR 24-MAY-2001; 2001US-0293589P.  
 PR 25-MAY-2001; 2001US-0293747P.  
 PR 29-MAY-2001; 2001US-0294110P.  
 PR 30-MAY-2001; 2001US-0294434P.  
 PR 14-AUG-2001; 2001US-0312182P.  
 PR 17-AUG-2001; 2001US-0313173P.  
 PR 17-AUG-2001; 2001US-0313187P.  
 PR 12-SEP-2001; 2001US-0318787P.  
 PR 12-SEP-2001; 2001US-0318744P.  
 PR 15-NOV-2001; 2001US-0335910P.  
 PR 28-NOV-2001; 2001US-0333881P.  
 PR 28-NOV-2001; 2001US-0333942P.  
 PR 03-JAN-2002; 2002US-0345716P.  
 PR 04-JAN-2002; 2002US-0345220P.  
 PR 01-MAY-2002; 2002US-00136071.  
 XX  
 XX (CURAGEN CORP.  
 XX  
 PI Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Casman SU,  
 PI Edinger SR, Ellerman K, Gangoli BA, Gerlach VL, Gorman L,  
 PI Gunther E, Herrman JL, Ji W, Lepley DW, Lewin DA, Li L,  
 PI McDougall JR, Malyanekar UM, Mezes PD, Padigan M, Patrujan M,  
 PI Peyman JA, Rastelli L, Rieger DK, Rothenberg ME, Sheno Sg,  
 PI Smltson G, Spytek KA, Stone DJ, Taupier RJ, Tchiernev VT,  
 PI Verne CM, Voss EZ, Zerhusen BD, Zhong H, Miller CB;  
 XX  
 DR WPI: 2003-111987/10.  
 DR P-PSDB; ABR96184.  
 XX  
 PT New NOVX polypeptides and polynucleotides useful for treating or  
 PT preventing e.g. cardiomyopathy, atherosclerosis, hypertension, congenital  
 PT heart defects, aortic stenosis, atrial septal defect, or arteriovenicular  
 PT canal defect.  
 PT  
 XX  
 PS Claim 20; Page 197; 491pp; English.  
 PS  
 XX ACF16993 to ACF17000 encode the human G protein-coupled receptor (GPCR)  
 CC proteins, designated NOVX proteins, given in ABR96130 to ABR96191. The  
 CC NOVX sequences can have cytosolic, cardiovascular, antiaslathmic,  
 CC immunosuppressive, anti-HIV (human immunodeficiency virus), hypotensive  
 CC and antiarteriosclerotic activities, and can be used in gene therapy.  
 CC NOVX polypeptides can be used for treating a syndrome associated with a  
 CC human disease such as a pathology associated with the polypeptide. NOVX  
 CC polypeptides, polynucleotides and antibodies can be used for treating or  
 CC preventing e.g. cardiomyopathy, atherosclerosis, hypertension, congenital  
 CC heart defects, aortic stenosis, atrial septal defect, arteriovenicular  
 CC canal defect, pulmonary stenosis, prostate cancer, uterine cancer,  
 CC neoplasm, graft versus host disease, acquired immunodeficiency syndrome  
 CC (AIDS), Crohn's disease, multiple sclerosis, or bronchial asthma. The  
 CC nucleic acid sequences may be used in chromosome mapping, identifying  
 CC individual from minute biological samples (tissue typing), and in  
 CC forensic identification of a biological sample. ACF17000 to ACF17117  
 CC represent PCR primers and probes for the NOVX sequences, which are used  
 CC in an example from the present invention  
 CC  
 XX Sequence 1583 BP; 272 A; 556 C; 490 G; 264 T; 0 U; 1 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1,39E-159 Length: 1583  
 Score: 1635.00 Matches: 325  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 97.38% Indels: 0

DB: 10 Gaps: 0

US-10-617-443B-2 (1-334) x ACP16993 (1-1583)

```

QY 10 GYLeuHISGlnLeuSerSerProArgTyrIlePheAsnPheIleAlaAspValValGlu 29
DB 580 GGCTTCACCGAGTGGACGCGCGGCTCAAGATTCACTTCCTGCTCAAGCTGGTGGAG 639
QY 30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
DB 640 AAGATGGACACGCGCTGGTCCACATAGAGCTTCTCTGAGACACCGCGTGTGGCCGC 699
QY 50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
DB 700 AACGTGCCCCCTGTCACGCGGTTCTGCTTCATCATGTCAGAGCCGCGCTGATCATCACCC 759
QY 70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuValGln 89
DB 760 AATGCCACAGTGGTGTCCAGCAACAGGCTGCCCCCGGACGACGACGCTCAAGTGCAG 819
QY 90 LeuGlnAsnGlyAspSerTyrGlnAlaThrIleLysAspIleAspLysSerAspIle 109
DB 820 CTACAGATGGGAGCTCTATGAGGCCACCATCAAGACATCGACAAGAGTGGACATT 879
QY 110 AlaThrIleLysIleHisProLysIleLysLeuProValLeuLeuGlnGlyHisSerAla 129
DB 880 GCCACATCAAGATCCATCCCAAGAAAAAGCTCTCTGTGTGTGTGTGTGTGTGTGTGTGT 939
QY 130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
DB 940 GACCTGCGGCTGGGAGATTGTGTGGCCATGGGAGTCTCTTCCGCTCAAGACACA 999
QY 150 ValThrThrGlyIleValSerThrAlaGlnArgGluIleValArgGluLeuGlnArgAsp 169
DB 1000 GTGACAAACGGGCACTGTCACACACTGCGCCAGCGGAGCGGAGCTGCGCTCCGAGAC 1059
QY 170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189
DB 1060 TCCGACATGGATCATCATCCAGACGATGCCATCATCATCAAGGAACTCCGGGAGACA 1119
QY 190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
DB 1120 CTGTGTAACCTGATGGCGAGTATGGCATCAACACGCTCAAGGTCACGCGTGGCATC 1179
QY 210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
DB 1180 TCCCTTGCATCCCTCAGACCGCATCACACGGTTCCTCAACAGGTTCCAAAGACAGCAG 1239
QY 230 IleLysAspTyrPheLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
DB 1240 ATCAAAAGCTGGAAGAGCGCTTCATGCGCATACGGATGGGAGCATCACACCAAGCTG 1299
QY 250 ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVal 269
DB 1300 GTGATAGAGTGAAGGCGCAGCAACCCGAGCTTCCCAAGAGTCAAGATTTATGTG 1359
QY 270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleIle 289
DB 1360 CAAGAGGTGGCGCAATTCACCTTCTCAGAGAGCGGCAATCCAAAGTGGTACATCATC 1419
QY 290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGlnValAlaLeuThr 309
DB 1420 GTCAAGGTCAACGGGCGTCTCTAGTGGACTCGAGTGGAGCGGCTGCTGACC 1479
QY 310 GluSerProLeuLeuGlnGluValArgArgGlyAsnAspAspLeuPheSerIleAla 329
DB 1480 GAGTCTCCTCTCTACTGAGAGTGGCGGCGGAGAACGACGACTCTCTTACGATCGCA 1539
QY 330 ProGluValValMet 334
DB 1540 CCTAGAGGTGTGATG 1554

```

AAD23855  
ID AAD23855 standard; cDNA; 1797 BP.  
XX  
AC AAD23855;  
XX  
DT 07-MAR-2002 (first entry)  
XX  
DE Human protease PRTS-14 cDNA.  
XX  
XX Human, protease, PRTS-14; tranquiliser; gene therapy; vaccine; allergy;  
KW infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;  
KW gastroenteritis; psoriasis; Alzheimer's disease; mental disorder; cancer;  
KW gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;  
KW epithelial disorder; uterine; anorexia; trauma; asthma; eczema; nausea;  
KW hypertension; neurological disorder; Parkinson's disease; drug screening;  
KW cardiast; cell proliferative disorder; multiple sclerosis; osteoporosis;  
KW diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;  
KW autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;  
KW developmental disorder; reproductive disorder; infertility; diarrhoea;  
KW dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FH CDS 181..1542  
FT /\*tag= a  
FT /product= "Human protease PRTS-14 protein"  
FT sig\_peptide 181..231  
FT /\*tag= b  
FT mat\_peptide 232..1539  
FT /\*tag= c  
FT /product= "Mature human protease PRTS-14 protein"  
XX  
PN MO200183775-A2.  
XX  
PD 08-NOV-2001.  
XX  
PF 04-MAY-2001; 2001WO-US014651.  
XX  
PR 04-MAY-2000; 2000US-0202082P.  
PR 11-MAY-2000; 2000US-0203566P.  
PR 17-MAY-2000; 2000US-0205803P.  
PR 25-MAY-2000; 2000US-0207477P.  
PR 01-JUN-2000; 2000US-0209402P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
XX Deleane AM, Lal P, Hafalia A, Patterson C, Walla NK, Kearney L;  
PI Tribouley CM, Khan FA, Yao MG, Baughn MR, Azimul Y, Elliott VS;  
PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DM;  
PI Reddy R, Yue H, Tang YT;  
XX  
DR WPI: 2002-034518/04.  
DR P-PDB; AAE4349.  
XX  
PT Novel human proteases and polynucleotides encoding the proteases, useful  
PT for treating, diagnosing or preventing cell proliferative,  
PT cardiovascular, autoimmune/inflammatory, neurological and developmental  
PT disorders.  
XX  
PS Claim 5; Page 150-151; 151pp; English.  
XX  
XX The invention relates to human proteases (PRTS-14) and its corresponding  
CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for  
CC the diagnosis, treatment and prevention of disorders associated with  
CC increased or decreased expression of PRTS. Examples of such disorders  
CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,  
CC hepatitis, psoriasis and cancer); autoimmune/inflammatory disorders  
CC (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,  
CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,  
CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and  
CC viral, bacterial, fungal, parasitic, protozoal and helminthic infections)  
CC ; cardiovascular disorders (myocardial infarction, ischaemic heart



CC 2814981. The gene, protein, and antibody sequences can be used in the  
 CC diagnosis, and treatment or prevention of a disease associated with its  
 CC altered expression. The diseases that can be treated are matrix-  
 CC remodeling diseases, including cancer, cardiomyopathy, arthritis,  
 CC angiogenesis, diabetic necrosis, atherosclerosis, fibrosis, and  
 CC ulceration

XX Sequence 2040 BP; 427 A; 635 C; 602 G; 376 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.:	1,96e-159	Length:	2040
Score:	1635.00	Matches:	325
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.38%	Indels:	0
DB:	3	Gaps:	0

US-10-617-443b-2 (1-334) x AA52362 (1-2040)

QY 10 GYLVEHIGLNULEUSERSERPROAGTYRFLYPHEAHPHEILAEALASPVALVALGILU 29  
 DB 72 GGTCTCCACCACTGAGCAGCCCGCTCAAGTTCAATTCTGACCGTGGTGAG 131  
 QY 30 LVEILAEALPROALVALVALHISILEGLULEUPHELEUARGHISPROLEUPHEGLYARG 49  
 DB 132 AAGATGCAACCAACCGGTGTCACATAGACCTTCTCTGAGACACCCGCTTTGGCCGC 191  
 QY 50 AENVALPROLEUSERSERGLYSERGLYPHEILEMERSERGLUAGLYLEULEILEIERNR 69  
 DB 192 AAGTCCCTCTGCGAGCGGTTCTGGCTTCATCATGTCAGAGCCGCGCTGATCATCACC 251  
 QY 70 AENALHIVAVLVALSERSERSENSEVALALAPROGLYARGINGLNULEULYVALGIN 89  
 DB 252 AATGCCACCGTGTGTCAGACCAAGTGTGCCCCGGGAGGAGCAAGCTCAAGGTGAG 311  
 QY 90 LEUGLAENGLYAPSERTRYGLUALATHTRIELYEAPLLEAPLYSLYSESERAPLLE 109  
 DB 312 CTACAGATGGAGCTCTCTATGAGGCCACATCAAGACATCAAGAAAGTGGACATT 371  
 QY 110 ALATHTRIELYSILEHISPROLYSLYSLYALEUPROVALLEULEULEUGLYHISSEVAL 129  
 DB 372 GCCACCATCAAGATCCATCCCAAGAAAGCTCCCTGTGTGTGCTGGGCTCATCGGCC 431  
 QY 130 ASPLLEUARGPROGLYGLUPHEVALVALHISILEGLYSEPROBHEALILEUGINAEHTH 149  
 DB 432 GACCTGGGCTGGGAGTTGTGTGGCCATCGCGCTCCCTCGCTCAAGAACACA 491  
 QY 150 VALTHTHTRGILYILEVALSERTHRALAGLARGGLUGLYARGGLULEUGLYLEUARGASP 169  
 DB 492 GTACACACGGGATCTGTCAACATCGCCACCGGAGGAGGAGGAGCTGGGCTCCGGGAC 551  
 QY 170 SERASPMERAPTRYILEGINTHRAAPALAILLELEANTYRGLYANSENSEGLYGLYPRO 189  
 DB 552 TCGACATGAGTACATCCAGACGAGATGCATCATCACTACCGGAACTCCGGGGGACCA 611  
 QY 190 LEUVALHLENLEUARGLYGLUVALILEGLYILEASHTHRIEULYSVALTHRALAGLYLE 209  
 DB 612 CTGTGTAACTGTGATGGCGAGGTCATTGCAACACGCTCAAGGTCACAGCGCTGGGATC 671  
 QY 210 SERPHEALATLEPROSEERAPARGLIETHTARGPHELEUTHRGUHPHEGINAPLYEGLN 229  
 DB 672 TCTCTTGCCATCCCTCTAGACCGCATACACGGTCTCTCAAGAGTTCCAAAGACACAG 731  
 QY 230 ILELYASPTTPLYSLYSLYARGPHEILEGLYILEARGWECARGTHRIETHRPROSERLEU 249  
 DB 732 ATCAAAAGCTGGAAGAGCCCTTCATCGCATACGATGCGAGATCAACACCAAGCTTG 791  
 QY 250 VALASPLULEULYSALASERSENPROASPHENPROGLUVALSERSERGLYILETYVAL 269  
 DB 792 GTGGATAGCTGAAGGCGACACCCGAGACTTCCCAAGAGGTCAGCAAGGAGATTATATG 851  
 QY 270 GLNGLUVALALPROASNSERPROSERGLNARGGLYGLYILEGINAEPGLYASPLILEIE 289

DB 852 CAAGAGTTGGCCGCAATTCACCTTCTCAGAGAGCGGCATCAAGATGTGATCATCATC 911  
 QY 290 VALLYVSLAENGLYARGPROLEUVALASPSERSERGLULEUGLNUALVALLEUHR 309  
 DB 912 GTCAAGGTCAACGGGCGTCTTACTAGACTTCGACTGAGCTGCAGAGCGCGTGTGACC 971  
 QY 310 GLUSERPROLEULEULEUGLNUVALARGARGLYAENAPAPLEULEUPHESERTLEALA 329  
 DB 972 GAGTCTCTCTCTCTATGAGAGGTGGCGCGGGAACGACGACTCTCTTACGATCCCA 1031

QY 330 PROGLUVALVALMET 334

DB 1032 CCTGAGGTGTCATG 1046

## RESULT 12

AA42458  
 AAL42458 standard; DNA; 2040 BP.

AC AAL42458;

XX 11-JUL-2002 (first entry)

DE Human matrix-remodeling-associated nucleotide 14.

KW Human; ds; matrix-remodeling gene; extracellular matrix; gene;

KW matrix-remodeling-associated nucleotide; screening;

KW matrix remodeling-associated disease; angiogenesis; arthritis;

KW atherosclerosis; cancer; cardiomyopathy; diabetic necrosis; fibrosis;

XX ulceration.

OS Homo sapiens.

XX US2002019000-A1.

PD 14-FEB-2002.

XX 26-MAR-2001; 2001US-00818143.

PF 09-OCT-1998; 98US-00169289.

PR (WALKER) WALKER M G.

PA (VOLK) VOLK MUTH W.

PA (KLIN/) KLINGLER T M.

XX Walker MG, Volkmut W, Klingler TM;

XX WPI; 2002-338319/37.

XX New isolated polynucleotide coexpressed with matrix-remodeling genes,

XX useful in diagnosis, prognosis, prevention and treatment of diseases

XX associated with matrix-remodeling such as angiogenesis, arthritis and

XX cancer.

XX Claim 2; Page 29-30; 63pp; English.

XX The invention comprises human nucleotide sequences which are co-expressed

XX with matrix-remodeling genes. Matrix-remodeling is associated with the

XX construction, destruction and reorganization of extracellular matrix

XX components. The matrix-remodeling-associated nucleotides of the invention

XX are useful for screening for and purifying ligands that specifically bind

XX to the nucleotides of the invention. The matrix-remodeling-associated

XX nucleotides of the invention are also useful in the diagnosis, prognosis,

XX prevention, treatment and evaluation of therapies for diseases associated

XX with matrix remodeling (e.g. angiogenesis, arthritis, atherosclerosis,

XX cancer, cardiomyopathy, diabetic necrosis, fibrosis and ulceration). The

XX present DNA sequence represents a human matrix-remodeling-associated

XX nucleotide of the invention

XX Sequence 2040 BP; 427 A; 635 C; 602 G; 376 T; 0 U; 0 Other;

XX Alignment Scores:

XX Pred. No.: 1,96e-159 Length: 2040

XX Score: 1635.00 Matches: 325

Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 97.38%  
 DB: 6  
 Gaps: 0

US-10-617-443B-2 (1-334) x AAL42458 (1-2040)

```

QY 10 GlyLeuHisGlnLeuSerSerProArgTyrIlePheAsnPhelLeuAspValValGlu 29
DB 72 GGTCTCCACAGCTAGACAGCCCGCTCAAGTTCACTTCTGAGCTGAGCTGGAG 131
QY 30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
DB 132 AAGATCGACAGCGCGTGTCCACATAGAGCTCTTCTGAGACACCCGCTGTTGGCCG 191
QY 50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGlyValAglyLeuIleIleThr 69
DB 192 AACGTGCCCTGTCCAGCGGTTCTGGCTTCATCATGTGAGAGCGCGCTCATATCCACC 251
QY 70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuValGln 89
DB 252 AATGCCACGTGTGTCCAGCAAGTCTGCCCCGAGGACAGCACTCAAGTGCAG 311
QY 90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleuAspIleAspGlySerSerAspIle 109
DB 312 CTACAGATGGGAGCTCTATAGAGCCACCATCAAGACATGACAAAGATCGGACATT 371
QY 110 AlaThrIleLysIleHisProLysIleValLeuProValLeuLeuGlyHisSerAla 129
DB 372 GCCACATCAAGATTCATCCCAAGAAAAAGCTCCCTGTGTGTGCTGGTCACTCGGCC 431
QY 130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
DB 432 GACCTGCGGCTGGGAGATTGTGTGGCCATCGGACATCCCTTCGCTCAGAACCA 491
QY 150 ValThrThrGlyIleValSerThrAlaGlnArgGluArgGluLeuArgAsp 169
DB 492 GTGACACGGGCACTGTGACAGCACTGCCAGGGGAGGAGGAGCTGGCTCCGGGAC 551
QY 170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyPro 189
DB 552 TCCCACTGAGCTCATCCAGACGGATCCATCATCACTCAAGGGAATCCGGGGGACCA 611
QY 190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuValThrAlaGlyIle 209
DB 612 CTGGTGAACCTGGAGTGGGAGGTCAATGGCAACAGCTCAAGGTCACGGCTGGCATC 671
QY 210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspGlyGln 229
DB 672 TCCTTTGGCATCCCTCAGACCGCATCAGCGTTCTCCACAGAGTTCCAGAGACAGCAG 731
QY 230 IleLysAspTyrPylsValArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
DB 732 ATCAAAAGCTGGAAGAGCGCTTCATCGGCATACGGAATGGCAGATCACCAAGCCTG 791
QY 250 ValAspGlyLeuLeuValAsnSerAspProAspPheProGluValSerSerGlyIleTyrVal 269
DB 792 GTGATGAGCTGAAGAGCGCAGAACCCGAGCTCCAGAGGTCACACAGTGAATTTATGTG 851
QY 270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleIle 289
DB 852 CAAGAGTGGCGCGGAATTCACCTTCAGAGAGCGGCGCATCCAAAGATGAGCATCATC 911
QY 290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluValValLeuThr 309
DB 912 GTCAAGGCTCAACGGGCGCTCTCTAGTGAAGTGAAGTGCAGAGAGCGCTGCTGACC 971
QY 310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspPheLeuPheSerIleAla 329
DB 972 GAGTCTCCTCTCTCACTGAGAGGTGGCGGGGAGACGACCTCTCTTCAGCATGCGCA 1031
QY 330 ProGluValValMet 334
  
```

DB 1032 CCTGAGGTGTCATG 1046

RESULT 13

ID ABEK51497 standard; CDNA; 2094 BP.

AC ABEK51497;

DT 13-AUG-2002 (first entry)

DE CDNA encoding human LP protein LP241.

XX Human; LP241; inflammatory disorder; diabetes; bone disease;

XX cardiovascular disease; male reproductive system disease; osteoporosis;

KW Paget's disease; myeloma; Alzheimer's disease; contraceptive;

KW liver cancer; growth factor-mediated disease; anophylaxis; coagulation;

XX sepsis; skeletal muscle dystrophy; asthma; breast cancer; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 206..1567

FT /tag= a

FT /product= "Secreted protein, LP241"

XX MO200232939-A2.

XX 25-APR-2002.

XX 10-OCT-2001; 2001WO-US027759.

XX 19-OCT-2000; 2000US-0241813P.

XX (EIL) LILLY & CO EIL.

XX Lu D, Song HY, Su EW, Wang H;

XX WPI: 2002-454591/48.

XX P-PSDB; AA097040.

XX New secreted human LP polypeptides or polynucleotides, useful for

PT treating mammals suffering from conditions associated with aberrant

PT levels of an LP polypeptide, e.g. cancers, osteoporosis, Paget's disease

PT or Alzheimer's disease.

XX Claim 1; Page 144-146; 148pp; English.

XX The invention relates to isolated human polypeptides designated LP102,

XX LP187, LP190 and LP241, and the polynucleotides encoding them. The LP

XX polypeptide or the LP polypeptide agonist is useful for treating a mammal

XX suffering from a disease, condition or disorder associated with aberrant

XX levels of an LP polypeptide. The LP polypeptides or polynucleotides, or

XX the antibodies are useful for treating, preventing or diagnosing cancers,

XX inflammatory disorders, bone diseases or cardiovascular diseases. In

XX particular, the LP102 polypeptides, polynucleotides or antibodies are

XX useful for diagnosing, preventing or treating male reproductive system

XX diseases, osteoporosis, Paget's disease, metabolic or myeloma associated

XX bone diseases or Alzheimer's disease. These are also useful as

XX contraceptive agents. The LP187 polypeptides, polynucleotides or

XX antibodies are particularly useful for diagnosing, preventing or treating

XX liver cancer, as well as other growth factor-mediated diseases and

XX conditions. The LP190 polypeptides, polynucleotides or antibodies are

XX useful for diagnosing, preventing or treating asthma, anaphylaxis, and

XX diseases related to coagulation or sepsis. LP241 polypeptides,

XX polynucleotides or antibodies are useful especially for diagnosing,

XX preventing or treating skeletal muscle dystrophy, breast cancer or

XX diabetes. The LP polynucleotides are also useful for constructing DNA

XX vectors that may be employed in medicine, or for preparing the proteins.

XX The present sequence represents the coding sequence of LP241

XX Sequence 2094 BP; 360 A; 737 C; 633 G; 364 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2, 03e-159	Length:	2094
Score:	1635.00	Matches:	325
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.38%	Indels:	0
DB:	6	Gaps:	0

US-10-617-443B-2 (1-334) X ABK51497 (1-2094)

[illegible]

QY		330	ProglutininMet	334
Db		1550	CCTGAGGTGCATG	1564
		RESULT 14		
		ACFI2810		
		ID ACFI2810	standard; DNA; 2543 BP.	
		XX AC	ACFI2810;	
		XX AC		
		DT 09-SEP-2003	(first entry)	
		DE		
		XX Human PRSP clone encoding sequence #1.		
		KW Endometrium; placenta; serine protease; gynecological; cytostatic;		
		KW cardiant; PRSP; infertility; endometriosis; cancer; pregnancy; human; ds.		
		OS Homo sapiens.		
		XX		
		XX Key Location/Qualifiers		
		FH FT 1..2543		
		FT CDS /tag= a		
		FT /product= "PRSP"		
		FN WO2003011905-A1.		
		PX		
		PD 13-FEB-2003.		
		PF 30-JUL-2002; 2002WO-AU001010.		
		PR 30-JUL-2001; 2001AU-00006707.		
		PA (PRIN-) PRINCE HENRY'S INST MEDICAL RES.		
		PI Nle G, Salomonsen LA, Li Y, Hampton AL, Findlay JK,		
		XX WPI; 2003-268108/26.		
		DR P-PDBS; ABR91169.		
		PT New nucleic acid encoding a protein having serine protease activity and		
		PT an insulin-like growth factor-binding motif, useful for preparing a		
		PT composition for treating a pregnancy-related serum protease-related		
		PT condition e.g., infertility.		
		PS Claim 5; Fig 3a; 156pp; English.		
		XX The present sequence relates to a new isolated nucleic acid molecule,		
		CC which is expressed in endometrium and placenta and is upregulated in		
		CC pregnant uterus and is highly expressed during placental development,		
		CC CC encodes a protein having serine protease activity and has an insulin-like		
		CC growth factor (IGF)-binding motif. The compound is considered		
		CC gynecological, cytostatic and cardiant. The enzyme is specifically		
		CC expressed in association with embryo implantation and placentaion in a		
		CC pregnant uterus. The nucleic acid is useful for preparing a composition		
		CC for treating PRSP-related condition e.g., infertility, endometriosis,		
		CC cancer or a disease of the heart, testis or ovaries. Further, it is		
		CC useful for detecting, diagnosing or monitoring a condition involving a		
		CC change in PRSP expression. The sequence is present in the exemplification		
		CC of the specification		
		XX Sequence 2543 BP; 458 A; 858 C; 777 G; 450 T; 0 U; 0 Other;		
		SQ		
		Alignment Scores:		
		Pred. No.: 2,65e-159	Length: 2543	
		Score: 1635.00	Matches: 325	
		Percent Similarity: 100.00%	Conservative: 0	
		Best Local Similarity: 100.00%	Mismatches: 0	
		Query Match: 97.38%	Indels: 0	
		DB: 8	Gaps: 0	
		US-10-617-443B-2 (1-334) x ACFI2810 (1-2543)		
QY	10 GlyLeuHisGlnLeuSerSerProArgTyrItyrPheAnpHeIIeAlaAspValValGlu 29			



```

Db      577 GGTCTCCACGACGAGGAGCCGGCTCAAGTTCACTTCATGTCGAGTGGAG 636
Qy      30  LysilleaProAlaValHisIleGluPheLeuArgHisProLeuPheIlyArg 49
Db      637 AAGATCGACACGCGGTGTCCATGAGCTCTTCCGAGACACCTCGTGTGGCGG 696
Qy      50  AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAglIleuIleIleThr 69
Db      697 AACGTGCCCCCTGTCCAGGGTTCTGCTTCATCATGTAGAGCGCGCTGATCATACC 756
Qy      70  AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuValGln 89
Db      757 AATGCCCACTGTGTGTCCAGCAACAGTGTCTCCCGGCGACGACAGCTCAAGGTGAG 816
Qy      90  LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysSerAspIle 109
Db      817 CTACAGAAATGGGAGCTCTATGAGCGCCACATCAAGACATCGACAAAGTCCGACATT 876
Qy      110 AlaThrIleLysIleHisProLysIleLysLeuProValLeuLeuLeuGlyHisSerAla 129
Db      877 GCCACCATCAAGATTCATCCCAAGAAAAGCTCCCTGTGTGTGTGTGTGTGTGTGTGT 936
Qy      130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
Db      937 GACCTGGCGGCTGGGAGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 996
Qy      150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyValGlyLeuArgAsp 169
Db      997 GTGCAACCGGGCATGTCTAGACATGCCACGCGGAGGCGAGGAGCTGGCGCTCCGGGAC 1056
Qy      170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyPro 189
Db      1057 TCCACATGGAATCAATCCACAGCGATGCCATCATCACTACGGAAGTCCGCGGAGACCA 1116
Qy      190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
Db      1117 CTGTGTAACTCGATGGAGGTGATTCATTCATCAACGCTCAAGGTCAAGCTGGGATC 1116
Qy      210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
Db      1177 TCCCTTGCACATCCCTCAGACCGCATCAACGCTTCTCAGAGGTTCCAGACAGACAG 1236
Qy      230 IleLysAspTyrLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
Db      1237 ATCAAAGACTGAGAGAGCGCTTCAATCGCATACGAGATGCGGAGATCAACCAAGCTG 1296
Qy      250 ValAspGluLeuLysValAspAsnProAspPheProGluValSerSerGlyIleTyrVal 269
Db      1297 GTGATGAGCTGAGAGGCGCAGAACCCCGGACTTCCAGAGGTCCAGAGTGAATTTATGTG 1356
Qy      270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleIle 289
Db      1357 CAGAGGCTGGCGCGAATTCCTTCTCAGAGAGCGGCATCCAAAGTGTGATCATCATC 1416
Qy      290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThr 309
Db      1417 GTCAAGGTCAACGGGCGCTCTAGTGTGATCTGAGTGAAGCTGAGAGGCGGTCTCAACC 1476
Qy      310 GluSerProLeuLeuLeuGluValArgArgLysAsnAspAspLeuLeuPheSerIleAla 329
Db      1477 GAGCTCTCTCTCTCACTGAGAGTCCGCGGGGGAACACACCTCTCTTCAAGATCGCA 1536
Qy      330 ProGluValValMet 334
Db      1537 CTGAGGTGTGATG 1551

```

RESULT 15  
 ADO01092  
 ID ADO01092 standard; DNA; 2554 BP.  
 AC ADO01092;  
 XX

```

DT      29-JUN-2004 (first entry)
XX      Human serine protease HTRA3 related DNA, SEQ ID 7.
DE      Cytostatic; Serine protease inhibitor; Apoptosis stimulator; human;
XX      serine protease; enzyme; HTRA3; pancreatic cancer; cancer; ds.
XX      Homo sapiens.
XX      WO2004039407-A1.
XX      13-MAY-2004.
XX      30-OCT-2003; 2003WO-JP013920.
XX      01-NOV-2002; 2002JP-00320075.
XX      27-JAN-2003; 2003JP-00017892.
XX      (TAKE ) TAKEDA CHEM IND LTD.
XX      Horikoshi K, Kitahara O, Matanabe T, Taniyama Y, Nishizawa S;
XX      WPI; 2004-400073/37.
XX      Apoptosis inducers inhibiting activity or expression of serine protease.
XX      HTRA3 for treatment, prevention and diagnosis of pancreatic cancer.
XX      Disclosure; SEQ ID NO 7; 124pp; Japanese.
XX      The present invention relates to apoptosis inducers and protease
XX      inhibitors, which contain a compound inhibiting the activity or
XX      expression of human serine protease HTRA3 (ADO01090 and ADO01115). The
XX      invention is useful for the treatment, prevention and diagnosis of
XX      pancreatic cancer, and also of other cancers including colon, breast,
XX      lung, prostate, oesophagus, stomach, liver, spleen, kidney, gall bladder,
XX      ovary, testis, bladder, thyroid and brain cancer and blood neoplasms. The
XX      present sequence was used to illustrate the invention.
SQ      Sequence 2554 BP; 447 A; 866 C; 786 G; 455 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.67e-159 Length: 2554
Score: 1635.00 Matches: 325
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.38% Indels: 0
DB: Gaps: 0
US-10-617-443B-2 (1-334) x ADO01092 (1-2554)
Qy      10  GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
Db      606 GGTCTCCACGACGAGTGCAGACCCGCGCTCAAGTTCACTTCATGTCGAGTGGAG 665
Qy      30  LysilleaProAlaValHisIleGluPheLeuArgHisProLeuPheIlyArg 49
Db      666 AAGATCGACACGCGGTGTCCATGAGCTCTTCTGAGACACCGCTGTGTGGCGG 725
Qy      50  AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAglIleuIleIleThr 69
Db      726 AACGTGCCCCCTGTCCAGGGTTCTGCTTCATCATGTAGAGCGCGCTGATCATACC 785
Qy      70  AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuValGln 89
Db      786 AATGCCCACTGTGTGTCCAGCAACAGTGTCTCCCGGCGACGACGCTCAAGGTGAG 845
Qy      90  LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysSerAspIle 109
Db      846 CTACAGAAATGGGAGCTCTATGAGCGCCACATCAAGACATCGACAAAGTCCGACATT 905
Qy      110 AlaThrIleLysIleHisProLysIleLysLeuProValLeuLeuLeuGlyHisSerAla 129
Db      906 GCCACCATCAAGATTCATCCCAAGAAAAGCTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGT 965

```

QY 130 AspleuAArgProGlyGluPheValAlaIleGlySerProPheAlaLeuGlnAsnThr 149  
 |||||  
 DB 966 GACCTGCGGCGCTGGGGAGTTTGTGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA 1025  
 QY 150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAsp 169  
 |||||  
 DB 1026 GTGACAAACGGGCGCTCTCAGCACTGCCGCGGAGGGCAGGGAGCTGGGCTCCGGGAGC 1085  
 QY 170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189  
 |||||  
 DB 1086 TCCGACATGACATCAATCCAGACGAGTCCATCATCACTACGGGAATCCGGGGAGACCA 1145  
 QY 190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209  
 |||||  
 DB 1146 CTGGTGAACCTGGATGGCGAGTCATTTGGCATCACACGCTCAAGGTCAAGCTGGCATC 1205  
 QY 210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229  
 |||||  
 DB 1206 TCCCTTTCCTCCATCCCTCAGACCGCATCACGCTTCTCTCAGAGTTCCAGACAAAGCAG 1265  
 QY 230 IleLysAspTyrPylsArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249  
 |||||  
 DB 1266 ATCAAAGACTGGAGAGAGCGCTTCATCGCATACGATGCGAGCATCACACCAAGCCTG 1325  
 QY 250 ValAspGluLeuLysValAsnAspProAspPheProGluValSerSerGlyIleTyrVal 269  
 |||||  
 DB 1326 GTGATGAGCTGAAGGCGCAGCACCCGGACTTCCAGAGGTCAAGTGAATTATGTG 1385  
 QY 270 GlnGluValAlaProAsnSerProSerGlnArgIleGlyIleGlnAspGlyAspIleIle 289  
 |||||  
 DB 1386 CAAGAGTTGGCGCAATTCACTTCTCAGAGAGCGGCATCCAAAGTGTGACATCATC 1445  
 QY 290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGlnAlaValLeuThr 309  
 |||||  
 DB 1446 GTCAAGGTCAACGGGCGTCTCTAGTGACTCGAGTGAGCTCGAGAGGCGGTGCTGACC 1505  
 QY 310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329  
 |||||  
 DB 1506 GAGTCTCCTCTCTCTGAGAGGTGGGGGGGAGACGACCTCTCTTCAGCATCGCA 1565  
 QY 330 ProGluValValMet 334  
 |||||  
 DB 1566 CCGAGGTGTGATG 1580

Search completed: June 29, 2005, 21:40:04  
 Job time : 956.732 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - nucleic search, using frame\_blue\_p2n model

Run on: June 29, 2005, 21:11:10 ; Search time 297.971 Seconds

(without alignments)  
1834.127 Million cell updates/sec

Title: US-10-617-443B-2

Perfect score: 1679

Sequence: 1 MHLLPASPAGHQLSPRYK.....LEVNRGNDLLFSLAPEVVM 334

Scoring table:  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 segs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cg2\_1/USPFO.spool/US10617443/runat\_29062005\_135932\_26096/app\_query.fasta\_1.718  
-DB=Issued\_Patents\_NA -QMT=faetap -SUFFIX=rml -MINMATCH=0.1 -LOPCL=0  
-LIST=45 -DOCL=1 -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-MODE=LOCAL -OUTPMT=p2n -NDB=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US10617443.qcch\_1.1\_93@runat\_29062005\_135932\_26096 -NCPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued\_Patents\_NA.\*

1: /cg2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cg2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cg2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cg2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cg2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cg2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1034.5	61.6	1521	4	US-09-949-016-4643
2	1034.5	61.6	2205	3	US-09-949-016-399
3	1034.5	61.6	2205	4	US-08-888-077A-41
4	1027.5	61.2	2036	3	US-08-923-454A-17
5	920	54.8	1859	3	US-09-724-864-22
6	882	52.5	1835	3	US-08-923-454A-7
7	882	52.5	2040	4	US-09-075-460-4
8	882	52.5	2187	3	US-08-923-454A-23
9	882	52.5	2187	3	US-08-923-454A-24
10	882	52.5	2476	3	US-09-008-271A-23
11	878	52.3	2187	3	US-08-923-454A-30
12	834.5	49.7	1787	3	US-08-923-454A-3

13	834.5	49.7	2551	3	US-08-923-454A-26	Sequence 26, Appl
14	779	46.4	1503	3	US-08-923-454A-5	Sequence 5, Appl
15	779	46.4	2144	3	US-08-923-454A-28	Sequence 28, Appl
16	579.5	34.5	539	1	US-08-322-742-18	Sequence 18, Appl
17	454	27.0	1230230	4	US-09-438-185A-1	Sequence 1, Appl
18	450.5	26.8	1425	4	US-09-489-039A-6523	Sequence 6523, Ap
19	444	26.4	1332	4	US-09-902-540-4672	Sequence 4672, Ap
20	444	26.4	24754	4	US-09-902-540-1230	Sequence 1230, Ap
21	441.5	26.3	1366	4	US-09-711-164-245	Sequence 245, App
22	437	26.0	1386	4	US-09-902-540-4598	Sequence 4598, Ap
23	437	26.0	27707	4	US-09-902-540-1226	Sequence 1226, Ap
24	431	25.7	1068	4	US-09-711-164-246	Sequence 246, App
25	430	25.6	1230025	4	US-09-198-452A-1	Sequence 1, Appl
26	428	25.5	1436	3	US-09-199-637A-131	Sequence 131, App
27	425	25.3	1428	4	US-09-252-991A-12923	Sequence 12923, A
28	425	25.3	1455	4	US-09-252-991A-12561	Sequence 12561, A
29	424	25.3	732	3	US-08-923-454A-1	Sequence 1, Appl
30	422	25.1	1980	2	US-08-350-741-1	Sequence 1, Appl
31	422	25.1	1980	2	US-08-463-875A-1	Sequence 1, Appl
32	418	24.9	1500	4	US-09-673-898-5	Sequence 5, Appl
33	417	24.8	1083	4	US-09-489-039A-6463	Sequence 6463, Ap
34	415	24.7	1110	4	US-09-673-898-7	Sequence 7, Appl
35	414	24.7	1242	4	US-09-388-090-5	Sequence 5, Appl
36	414	24.7	1326	4	US-09-388-089B-13	Sequence 13, Appl
37	414	24.7	1395	4	US-09-388-089B-10	Sequence 10, Appl
38	414	24.7	1395	4	US-09-388-090-3	Sequence 3, Appl
39	414	24.7	1611	1	US-08-485-569-1	Sequence 1, Appl
40	414	24.7	1611	1	US-08-480-993-1	Sequence 1, Appl
41	414	24.7	1611	2	US-07-903-079B-1	Sequence 1, Appl
42	413	24.6	1830121	4	US-09-557-884-1	Sequence 1, Appl
43	413	24.6	1830121	4	US-09-643-990A-1	Sequence 1, Appl
44	411.5	24.5	1488	4	US-09-489-039A-5255	Sequence 5255, Ap
45	411.5	24.5	1494	4	US-09-489-039A-5193	Sequence 5193, Ap

## ALIGNMENTS

RESULT 1  
US-09-949-016-4643  
Sequence 4643, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastsEQ for Windows Version 4.0  
SEQ ID NO 4643  
LENGTH: 1521  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-4643  
Alignment Scores:  
Pred. No.: 1.58e-120  
Score: 1034.50  
Percent Similarity: 80.43%  
Best local Similarity: 62.08%  
Query Match: 61.61%  
DB: 4  
Gaps: 2  
US-10-617-443B-2 (1-334) x US-09-949-016-4643 (1-1521)  
Cy 10 G1yueh1e1n1euser1er1ro1g1Ty1Ly1Phe1an1he1l1a1a1p1a1l1a1G1u 29

```

Db      2 GGGCAGAGATCCCAAGAGTTGGCCCATTAATATACTTTACGGCGACGGTGGAG 61
Qy      30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
Db      62 AAGATGGCCCTCCCGGCTTCATTCGAATTTGTTGCAAGGCTTCGTTTCTTAACGA 121
Qy      50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluIleGlyLeuIleIleThr 69
Db      122 GAGGTGCGGCTGCTAGTGGGTCTGGGTTTATTGTGCGAAGATGGACTGATCGGACA 181
Qy      70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyValArgIleGluLeuValGln 89
Db      182 AATGCCACGGTGTGACCAAC-----AAGACCGGGCTCAAAAGTTGAG 223
Qy      90 LeuGlnAsnGlyAspSerTyrgIuAlaThrIleLeuAspIleAspIleValSerAspIle 109
Db      224 CTGAAAGACGGTCCCACTTACGAAGCCAAATCAAGGATGTGATGAGAAAGCAGACATC 283
Qy      110 AlaThrIleLeuValIleHisProLysIleValLeuProValLeuLeuGluHisSerAla 129
Db      284 GCACTCATCAAAATTAACCAACGAGCGAAGCTCTGCTCTGCTGCTGGCCGCTCTCA 343
Qy      130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
Db      344 GAGCTCGGCGCGGAGAGTTCTGTGTGCGCATCGAAAGCCCGTTTCCCTTCAAAACAC 403
Qy      150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyValArgIleGluLeuArgAsp 169
Db      404 GTCAACACCGGGATCTGTGACACACCGCGAGCGAGCGGCGGCGGCGGCGCTCCCAAC 463
Qy      170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189
Db      464 TCAGACATGACTACATCCAGACCGAGCCCATCATCACTATGAAACCTCGGAGGCGCG 523
Qy      190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuValThrAlaGlyIle 209
Db      524 TTAGTAACTCGGACGGTGAAGATTGAATTAACTTTGAAATGACAGCTGGAATC 583
Qy      210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspGln 229
Db      584 TCTTTTGCAATCCCATCTGATTAAGATTAAAGTTCTCAACGAGTCCCATGCCACGAC 643
Qy      230 IleValAsp-----TrrPylsValArgPheIleGlyIleArgMetArgThrIleThr 246
Db      644 GCCAAAGGAAAGCCATCACCAAGAAAGATATATGATCCGAATGATGTCACCTCAG 703
Qy      247 ProSerLeuValAspGluLeuValAsnAsnProAspPheProGluValSerSerGly 266
Db      704 TCAGCAAGAACCAAGAGCTGAAGACCGGACCGGACCTTCCCAACGATCTCAGGA 763
Qy      267 IleTyrValGlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGly 286
Db      764 GCGTATATATTAAGATAATTTCTGTATACCCCGAGAGCTGTGTGTCTCAAGGAAAC 823
Qy      287 AspIleIleValIleValAsnGlyArgProLeuValAspSerSerGluLeuGlnAla 306
Db      824 GAGGTATATATGACATCAATGACAGCTCGGTGCTCCGCCATATATGACGCGCTC 883
Qy      307 ValLeuThrGluSerProLeuLeuGluValArgArgGlyAsnAspAspLeuPhe 326
Db      884 ATTAAAGGAAAGCAACCTGTAACATGTGTGTCGAGGGGTATGAATATCATGATC 943
Qy      327 SerIleAlaProGluValVal 333
Db      944 AAGATGATTCGCAAGAAATT 964

```

RESULT 2  
US-09-949-016-399

; Sequence 399, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

```

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 399
; LENGTH: 2036
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-399

```

## Alignment Scores:

```

Pred. No.: 2,576-120 Length: 2036
Score: 1034.50 Matches: 206
Percent Similarity: 80.18% Conservative: 61
Best Local Similarity: 61.86% Mismatches: 57
Query Match: 61.61% Indels: 10
Gaps: 2

```

US-10-617-443b-2 (1-334) x US-09-949-016-399 (1-2036)

```

Qy      4 AlaLeuProAlaSerAlaGlyLeuHisGluLeuSerProArgTyrIlePheAsnPhe 23
Db      503 GCGGAGCCTGCGGCGCAAGGCGAG-GAAGATCCCAACATTTGGCCGCAATTAATACCTT 561
Qy      24 IleAlaAspValValGluValIleAlaProAlaValValHisIleGluLeuPheLeuArg 43
Db      562 ATCGGCGAGTGTGTGAAGAAGATCCCGCTCGGTTCATTCGAATTTGTCGCAAG 621
Qy      44 HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGlu 63
Db      622 CTTCGGTTTTCTAAACGAGAGTGGGTGGTCTGAGTGTGATTTATGTGCGAA 681
Qy      64 AlaGlyLeuIleIleThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyArg 83
Db      682 GATGACATGATCTGACAAATGCCACGCTGTGACCAAC-----AAG 723
Qy      84 GlnGluLeuValGlnLeuGlnAsnGlyAspSerTyrgIuAlaThrIleLeuAspIle 103
Db      724 CACCGGTCCAAAAGTTGAGCTGAAGAACGGTCCCACTTACGAAGCCAAATCAAGGATGTG 783
Qy      104 AspIleValSerAspIleAlaThrIleValIleHisProLysIleValLeuProValLeu 123
Db      784 GATGGAAGAACGACATGCACTCATCAAAATTAACCAACGAGCGCAAGCTGCTGCTG 843
Qy      124 LeuLeuGlyHisSerAlaAspLeuArgProGlyGluPheValValAlaIleGlySerPro 143
Db      844 CTGCTGGCCCGCTCTCTCAGAGTGGCGCGGGAAGTTCTGTGTGCGCATGGAAGCCCG 903
Qy      144 PheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArgGluGlyArg 163
Db      904 TTTTCCCTTCAAAACACAGTCAACCCGCGGATCTGAGACACACCGACGCGCGGCAAA 963
Qy      164 GluLeuGlyLeuValArgAspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyr 183
Db      964 GAGCTGGGCTCCGCAACCTCAAGCTGACATCACTCAACGACGAGCGCATATCAACTAT 1023
Qy      184 GlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeu 203
Db      1024 GGAATCTGGAGAGCGCCCTTATGAACCTGGAACGTTGAAGTATTAACCTTTG 1083
Qy      204 LysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThr 223
Db      1084 AAGTGAACGCTGGAATCTCTTTGCAATCCCATCTGATTAAGATTAAAGTTCTCTCAG 1143

```

QY 224 GluPheGlnAspIleValGlnIleValAsp-----TrpIleValArgPheIleGlyIle 240  
Db 1144 GAGTCCCATGATCTCAAGAGCCGTAATATTAAGTAATCTTCTGATACCCGACGAAAGCT 1203  
QY 241 ArgMetArgThrIleThrProSerLeuValAspGluLeuValAspSerProAspPhe 260  
Db 1204 CGAATGATCTCAAGAGCCGTAATATTAAGTAATCTTCTGATACCCGACGAAAGCT 1263  
QY 261 ProGluValSerSerGlyIleTyrValGlnIleValAspProAspSerProSerGlnArg 280  
Db 1264 CCAGACGTCATCTCAAGAGCCGTAATATTAAGTAATCTTCTGATACCCGACGAAAGCT 1323  
QY 281 GlyGlyIleGlnAspGlyAspIleIleValIleValAsnGlyArgProLeuValAspSer 300  
Db 1324 GGTGTCCTCAAGAAACGACGTCATTAATCAAGTCATGACAGTCGTCGTCGTCGTCGTC 1383  
QY 301 SerGluLeuGlnIleValIleValLeuThrGluSerProLeuLeuGluValIleArgGly 320  
Db 1384 AATGATCTCAAGAGCCGTCATTAATTAAGGAAAGCAAGCCGTCATGTCGTCGTCGTCGTC 1443  
QY 321 AsnAspAspLeuLeuPheSerIleAlaProGluValVal 333  
Db 1444 AATGATCTCAAGAGTCATCAAGTCATTAATTAAGGAAAGCT 1482

RESULT 3

US-08-888-077A-41  
Sequence 41: Application US/08888077A  
Patent No. 602043  
GENERAL INFORMATION:  
APPLICANT: ST. GEORGE-HYSLUP, PETER H  
APPLICANT: ROMENS, JOHANNA M  
APPLICANT: FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE AND USES THEREFOR.  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LERNER, DAVID, LITTEMBERG, KUMHOLZ & MENTLIK  
STREET: 600 SOUTH AVENUE WEST  
CITY: WESTFIELD  
STATE: NJ  
COUNTRY: USA  
ZIP: 07090-1497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/888, 077A  
FILING DATE: 03-JUL-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/592,541  
FILING DATE: 26-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: PALIST, THOMAS M  
REGISTRATION NUMBER: 36,629  
REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP IV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 654-5000  
TELEFAX: (908) 654-7866  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2205 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..2205  
OTHER INFORMATION: /note="mutTMI-TM2"  
US-08-888-077A-41

Alignment Scores:  
Pred. No.: 2,93e-120 Length: 2205  
Score: 1034.50 Matches: 206  
Percent Similarity: 80.18% Conservative: 61  
Best Local Similarity: 61.86% Mismatch: 57  
Query Match: 61.61% Indels: 10  
Gaps: 2

US-10-617-443b-2 (1-334) X US-08-888-077A-41 (1-2205)

QY 4 AlaLeuProAlaSerAlaGlyLeuHisGlnLeuSerSerProArgTyrIlePheAspPhe 23  
Db 579 GCGGAGCTCTCGGCGCAAGGCGCAG-GAAGATCCCAACAGTTTGGCCCTTAATATTAATCTT 637  
QY 24 IleAlaAspValValGluValIleValProAlaValValHisIleGluLeuPheLeuArg 43  
Db 638 ATCCGAGCTGTGGGAAAGATGCCCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 697  
QY 44 HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGlu 63  
Db 698 CTTCGCTTTCTAAACGAGAGTCGCCGTGCTAGTGGCTGTGCTGTGCTGTGCTGTGCTGTGCT 757  
QY 64 AlaGlyLeuIleIleThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyArg 83  
Db 758 GATGACTGATCTGCAAAATGCCACGTGTGACCAAC-----AAG 799  
QY 84 GlnGlnLeuValGlnLeuGlnAsnGlyAspSerTyrGlnAlaThrIleValAspIle 103  
Db 800 CACCGGCTCAAGTGTGAGAAAGCGTCCACCTTAACGAAACCAAAATCAAGAGATGTG 859  
QY 104 AspIleSerSerAspIleAlaThrIleValIleHisIleProValValValLeuProValLeu 123  
Db 860 GATGAGAAACAGACATCGACATCATTAATTAATGACCAAGGAGGAGGAGGAGGAGGAGGAGG 919  
QY 124 LeuLeuGlyHisSerAlaAspLeuArgProGlyLeuPheValValAlaIleGlySerPro 143  
Db 920 CTGCTGCGCGCTCTCAAGAGTCGCGCGCGGAGAGTTCGTGTCGTCGTCGTCGTCGTCGTCGTC 979  
QY 144 PheAlaLeuGlnAsnThrValIleThrGlyIleValSerThrAlaGlnArgGluGlyArg 163  
Db 980 TTTTCCCTTAAACACAGTCACACCGGATCGTGACACACCCGAGGAGGAGGAGGAGGAGGAGG 1039  
QY 164 GluLeuGlyLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIleAsnTyr 183  
Db 1040 GAGCTGGGCTCCGCAATCAGACATGACTTACATCCAGCCGACATCACTAT 1099  
QY 184 GlyAsnSerGlyGlyProLeuValAsnLeuAspGlyValIleGlyIleAsnThrLeu 203  
Db 1100 GGAACCTCGGAGGCGCGTGAATTAACCTGACGAGTGAATGATTAACACTTTC 1159  
QY 204 LysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThr 223  
Db 1160 AAGTGACACTGGAATCTCTTTCGATCCCATCGATTAAGATTAAAGTTGCTTCACG 1219  
QY 224 GluPheGlnAspIleValGlnIleValAsp-----TrpIleValArgPheIleGlyIle 240  
Db 1220 GAGTCCCATGATCTCAAGAGCCGTAATATTAAGTAATCTTCTGATACCCGACGAAAGCT 1279  
QY 241 ArgMetArgThrIleThrProSerLeuValAspGluLeuValAspSerProAspPhe 260  
Db 1280 CGAATGATCTCAAGAGCCGTCATTAATTAAGTAATCTTCTGATACCCGACGAAAGCT 1339  
QY 261 ProGluValSerSerGlyIleTyrValGlnIleValAspProAspSerProSerGlnArg 280  
Db 1340 CCAGACGTCATCTCAAGAGCCGTCATTAATTAAGTAATCTTCTGATACCCGACGAAAGCT 1399  
QY 281 GlyGlyIleGlnAspGlyAspIleIleValIleValAsnGlyArgProLeuValAspSer 300  
Db 1400 GGTGTCCTCAAGAAACGACGTCATTAATTAAGTAATCTTCTGATACCCGACGAAAGCT 1459  
QY 301 SerGluLeuGlnIleValIleValLeuThrGluSerProLeuLeuGluValIleArgGly 320  
Db 1460 AATGATCTCAAGAGTCATTAATTAAGGAAAGCAAGCCGTCATGTCGTCGTCGTCGTCGTC 1519

QY 321 AenaspApplLeuPheSerIleAlaProGluVal 333  
DB 1520 AATGAAGATATCATGATCACAGTGAATTCGGAAGAAATT 1558

## RESULT 4

US-08-923-454A-17  
Sequence 17, Application US/08923454A  
Patent No. 6004794  
GENERAL INFORMATION:  
APPLICANT: Creasy, Caretha  
APPLICANT: Livi, George  
APPLICANT: Karan, Eric  
APPLICANT: Clinkenbeard, Helen  
APPLICANT: Browne, Michael  
APPLICANT: Southan, Christopher  
TITLE OF INVENTION: HUMAN SERINE PROTEASE  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/923,454A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/025436  
FILING DATE: 06-SEPT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: P50547  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5096  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2036 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORIGIN: Feature polymorphism at 1325  
US-08-923-454A-17

Alignment Scores:  
Pred. No.: 1 98e-119 Length: 2036  
Score: 1027.50 Matches: 205  
Percent Similarity: 79.88% Conservative: 61  
Best Local Similarity: 61.56% Mismatches: 58  
Query Match: 61.20% Indels: 10  
Gaps: 2

US-10-617-443B-2 (1-334) x US-08-923-454A-17 (1-2036)

QY 4 AlaLeuProAlaSerAlaGlyLeuHisGlnLeuSerProArgTyrLysPheAsn 23  
DB 503 GCGAGACCTGCGGCCCAAGGCGAG-GAAGATCCCAACAGTTTGGCCATTAATTAATTCTTT 561

QY 24 IleAlaAPValValGluLysIleAlaProAlaValHisIleGluLeuPheLeuArg 43  
DB 562 ATGCGGACGTGTGGAGAGATCCCTCCGCTGCTCATATGCAATTTGTTGGCAAG 621  
QY 44 HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleLeuSerGlu 63  
DB 622 CTTCCTTTCTTAAACGAGAGGTCCGGGTGCTAGTGGGTCTGGGTATTATGTGCGGAA 681  
QY 64 AlaGlyLeuIleIleThrAsnAlaHisValAlaSerSerAsnSerAlaAlaProGlyArg 83  
DB 682 GATGACCTGATCGTCAAAATGCCACGTGTGACCAAC-----AAG 723  
QY 84 GlnGlnLeuLysValGlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIle 103  
DB 724 CACCGGCTCAAGTTGAGCTGAAGAACGGTCCACTTTCGAAAGCCAAATCAAGATGTG 783  
QY 104 AspLysLysSerAspIleAlaThrIleLysIleHisProLysLysLeuProValLeu 123  
DB 784 GATGAGAAAGCAGACATCGACATCAAAATTGACCAACGAGGCAAGCTGCTGCTG 843  
QY 124 LeuLeuGlyHisSerAlaAspLeuArgProGlyGluPheValValAlaIleGlySerPro 143  
DB 844 CTGCTTGGCCGCTCCTCAGAGCTGCGCGGAGAGTTCGTGTGTCATCGGAAGCCCG 903  
QY 144 PheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArgGluGlyArg 163  
DB 904 TTTCCTTTCAAAACACAGTCCACCGGATCGTACACACCCAGGCGGCAAA 963  
QY 164 GlnLeuGlyLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIleLeuAsnTyr 183  
DB 964 GAGCTGGGGCTCCGCACTCAGACATGAGACTACATCCAGACCGCATCATCAAT 1023  
QY 184 GlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeu 203  
DB 1024 GGAACCTGGGAGCCCGCTTGAATACTGACGCTGAAGTGAATTAACCTTGTG 1083  
QY 204 LysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThr 223  
DB 1084 AATGTACAGCTGGAATCTCTTTCATCCATCTGTATTAATTAATAAGTTCTCTCAG 1143  
QY 224 GluPheGlnAspLysGlnIleLysAsp-----TPLYSLYsArgPheIleGlyIle 240  
DB 1144 GAGTCCCATGACCGACAGGCCAAAGAAAGCCATCACCAAGAAAGATATGTGTATC 1203  
QY 241 ArgMetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAsnProAspPhe 260  
DB 1204 CGAATGATGATCACTACCTCCGCAAGCCAAAGCGTGAAGGAGCCGCGGACTTC 1263  
QY 261 ProGluValSerSerGlyIleTyrValGlnGluValAlaProAsnSerProSerGlnArg 280  
DB 1264 CCGACGTGATCTCAGAGCGGTATATTAATTGAAGTATCTTGATATCCCCAGCAAGACT 1323  
QY 281 GlyGlyIleGlnAspGlyAspIleIleValLysValAsnGlyArgProLeuValAspSer 300  
DB 1324 GKTGCTCAAGAAAGAGAGCTATATATGATCATGACATCAATGACATCGTGGTCTCCGC 1383  
QY 301 SerGluLeuGlnGluAlaValLeuThrGluSerProLeuLeuLeuGluValArgArgGly 320  
DB 1384 AATGATGTACGACGATCAATTAAAGGAAAGCAACCTTGAACTGTGTGTGCGAGGGGT 1443  
QY 321 AenaspApplLeuPheSerIleAlaProGluVal 333  
DB 1444 AATGAAGATATCATGATCACAGTGAATTCGGAAGAAATT 1482

RESULT 5  
US-09-724-864-22  
Sequence 22, Application US/09724864  
GENERAL INFORMATION:  
APPLICANT: Watson, James D  
APPLICANT: Watson, James G.  
TITLE OF INVENTION: Polynucleotides, polypeptides expressed  
by the polynucleotides and methods for their use.

```

QY      192 nleuasp6g1yctluva11leegly1leanthr1euly5val1mr1a1agly1leser1pea1 212
          |||||::|||

```

LENGTH: 1835 base pairs

/ TYPE: nucleic acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: cDNA  
 / HYPOTHETICAL: NO  
 / ANTI-SENSE: NO  
 / FRAGMENT TYPE:  
 / ORIGINAL SOURCE:  
 / FEATURE:  
 / NAME/KEY: Coding Sequence  
 / LOCATION: 251...1624  
 / OTHER INFORMATION:  
 / US-08-923-454A-7

## Alignment Scores:

Pred. No.:	4,43e-101	Length:	1835
Score:	882.00	Matches:	169
Percent Similarity:	74.84%	Conservative:	72
Best Local Similarity:	52.48%	Mismatches:	71
Query Match:	52.53%	Indels:	10
DB:	3	Gaps:	2

US-10-617-443B-2 (1-334) x US-08-923-454A-7 (1-1835)

```

QY 15 SerSerProArgTyrIysPheAenPheIleAlaSPValValGluIysIleAlaProAla 34
DB 671 GCTTCTCCCGGAGCTCACTCACTTCATGCGAGATGTGTGAGAGACAGACCTGCC 730
QY 35 ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgHisValProLeuSer 54
DB 731 GTGGCTAATATCGAGATCTCGACCGGACCTTTCTGGGCGGAGGCTCTATCTCG 790
QY 55 SerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleLeuHisValVal 74
DB 791 AACGCTCAGGATTCGTGTGCTGCCGATGCGCTCATTTGCCAACCCCATGTGTG 850
QY 75 SerSerAenSerAlaAlaProGlyArgGlnGlnLeuValGlnLeuGlnAenGlyAap 94
DB 851 GCTGAT-----CGGCGAGAGTCCGTGTGAGACTCTCAACCGGCGAC 892
QY 95 SerTyrGlnAlaThrIleLeuAapIleAapIleAapIleAapIleAapIleAapIle 114
DB 893 ACCTATGAGGCCGTGTCACAGCTGTGATCCCGTGACAGACATCCCAACGCTGAGATT 952
QY 115 HisProLeuValLeuPheProValLeuLeuGlyHisSerAlaProLeuArgProGly 134
DB 953 CAGACTAAGAGGCTCTCCCGACGCTCTGGAGCGCTCAGCTGATGTCCGGCAAGG 1012
QY 135 GluPheValValAlaIleGlySerProPheAlaLeuGlnAenThrValThrThrGlyIle 154
DB 1013 GAGTTGTGTGTCCTCGGAGACTCCCTTGCACCTGCAACACGATCATCCGCGATT 1072
QY 155 ValSerThrAlaGlnArgGluArgGluLeuGlyLeuArgAapSerAapMetAapIyr 174
DB 1073 GTTAGCTCTGCTCAGCGTCAGCAGACAGACTGCGATCCCGCAACCAATGAGAAATAC 1132
QY 175 TleGlnThrAapAlaIleIleAenYrGlyAenSerGlyGlyProLeuValAenLeuAap 194
DB 1133 ATTCAAACGATGACCTATGATTTTGAAGACTTGGAGGTCCCTGTGTAACTCGGAT 1192
QY 195 GlyGluValIleGlyIleAenThrLeuValThrAlaGlyIleSerPheAlaIlePro 214
DB 1193 GGGAGGTATTTGATGATGAACACGATGAGTCAAGCTGGAATCTCTTGGCATCCCT 1252
QY 215 SerAapArgIleThrArgPheLeuThrGluPheGlnAapIyrGln-----Ile 230
DB 1253 TCTGATCGCTTCGAGAGTTTCTGCATCGTGGGGAAGAAAGAAATTCCTCTCCGGAATC 1312
QY 231 LysAapThrPylLeuValArgPheIleGlyIleArgMetArgThrIleThrProSerLeuVal 250
DB 1313 AGTGGGTCCAGAGGGGCTACATTTGGGTGATGATCTGACCTTGATCCCAAGCATCCCT 1372
QY 251 AapGluLeuValAalSerAenProAapPheProGluValSerSerGlyIleThrValGln 270
  
```

```

DB 1373 GCTGAACCTACAGCTTCGAGAACCAAGCTTTCCCATGTTCAAGCATGCTGTACTATCAT 1432
QY 271 GluValAlaProAenSerProSerGlnArgGlyIleGlnAapGlyAapIleIleVal 290
DB 1433 AAGTCATCTCGGCTCCCTGCGACACCGGCTGTGTGCGGCGCTGGATGTGATTTTG 1492
QY 291 LysValAenGlyArgProLeuValAapSerSerGluLeuGlnGlnAlaValLeuThrGlu 310
DB 1493 GCCATTGGGAGACAGATGCTCAAAATGCTGAAGATGTTATGAAAGCTTTCGAAACCCA 1552
QY 311 SerProLeuLeuGluValAlaArgArgGlyAenAapAapLeuLeuPheSerIleAlaPro 330
DB 1553 TCCAGTTGGGAGCTCAGATCCGCGGAGAGAAACACTGACTTATATGTACCCCT 1612
QY 331 GluVal 332
DB 1613 GAGTTC 1618

RESULT 7
US-09-075-460-4
/ Sequence 4, Application US/09075460A
/ Patent No. 6489136
/ GENERAL INFORMATION:
/ APPLICANT: Zervos, Anconia S.
/ TITLE OF INVENTION: CELL PROLIFERATION RELATED GENES
/ FILE REFERENCE: 10284/004001
/ CURRENT APPLICATION NUMBER: US/09/075,460A
/ EARLIER FILING DATE: 1998-05-08
/ EARLIER APPLICATION NUMBER: US 60/046,077
/ NUMBER OF SEQ ID NOS: 16
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 4
/ LENGTH: 2040
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ NAME/KEY: CDS
/ LOCATION: (248) ... (1834)
US-09-075-460-4

Alignment Scores:
Pred. No.: 5,28e-101 Length: 2040
Score: 882.00 Matches: 169
Percent Similarity: 74.84% Conservative: 72
Best Local Similarity: 52.48% Mismatches: 71
Query Match: 52.53% Indels: 10
DB: 4 Gaps: 2

US-10-617-443B-2 (1-334) x US-09-075-460-4 (1-2040)
QY 15 SerSerProArgTyrIysPheAenPheIleAlaSPValValGluIysIleAlaProAla 34
DB 881 GCTTCTCCCGGAGCTCACTCACTTCATGCGAGATGTGTGAGAGACAGACCTGCC 940
QY 35 ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgHisValProLeuSer 54
DB 941 GTGGCTAATATCGAGATCTCGACCGGACCTTTCTGGGCGGAGGCTCTATCTCG 1000
QY 55 SerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThrAenAlaHisValVal 74
DB 1001 AACGCTCAGGATTCGTGTGCTGCCGATGCGCTCATTTGCAACACGCGCATGTGCTG 1060
QY 75 SerSerAenSerAlaAlaProGlyArgGlnGlnLeuValGlnLeuGlnAenGlyAap 94
DB 1061 GCTGAT-----CGGCGAGAGTCCGTGTGAGACTGCTAAGCGGCGAC 1102
QY 95 SerTyrGlnAlaThrIleLeuAapIleAapIleAapIleAapIleAapIleAapIle 114
DB 1103 ACGATAGAGCGGTGTGTCACAGCTGTGATCCCGTGACAGACATCCGCTGAGGATT 1162
QY 115 HisProLeuValLeuPheProValLeuLeuGlyHisSerAlaProLeuArgProGly 134
  
```

```

Db      1163 CAGACTAAGGAGCCTCTCCCAACGCTGCTGGAGACCTGAGCTGATGTCGGGCAAGG 1222
Qy      135 GUPHPEVALVALAIAIEGLISERPHHEALALEUGINAMTHVALTHTHGLYILE 154
Db      1223 GAGTTTGTGTGGCCATGGGAAGTCCCTTGGACGTGCAAGAACACATACATCCGGCATT 1282
Qy      155 VALSERTHRAGLAINARGLUGLYARGGLUBENGLYLEUARGPESERAPMETAPTFR 174
Db      1283 GTTAGCTGTGCTCAGCGCTCCAGCCAGACCTGGAGACTCCCAACCAATGTGAATAC 1342
Qy      175 ILEGINTHRAPALAIIEIEASNTYRGLYANSENGLYLPROLEUVALASPLEUASP 194
Db      1343 ATTCAACTGATGACGATATTGATTGTGAAGAACTGTGAAGGTCCTGTTAACTGGAT 1402
Qy      195 GLYGLUVALIIEGLYILEASNTHEULYSVALTHRALAGLYILESERPHEALALEPRO 214
Db      1403 GGGAGGATGATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1462
Qy      215 SERAPARGILETHRARGPHELEUTHRGUPHEGLINAPLYSGIN-----ILE 230
Db      1463 TCTATGCTCTTCCAGAGCTTCTGATGCTGGGAAAAAGAAATTCCTCCCGAATC 1522
Qy      231 LYSAPTRPLYSLYSARGPHEIIEGLYILEARGMETARGTHRIETHRPROSERLEUVAL 250
Db      1523 AGTGGTCCCAAGCGGCTCATGATGGGGTGAATGATGATGATGATGATGATGATGATGAT 1582
Qy      251 ASPLULEULYSALASERAPHEPHEPROGLIUALISERSEGLYILETRYVALGIN 270
Db      1583 GCTAACCTACAGCTTCGAGAACCAAGCTTCCCATGATGTCAGCATGGTGTCTCATCAT 1642
Qy      271 GLUVALIAPROANSEPRSERSEGLINARGLYILEGINAPGLYASPLIIELEVAL 290
Db      1643 AAATCATCTCGGCTCCCTGACACCGGCTGCTGCGGCTGATGATGATGATGATGATGAT 1702
Qy      291 LYSVALANSGLYARGPROLEUVALASPSESEGLYINGLUALAVALIETHRGU 310
Db      1703 GCATTTGGGAGCAGATGATCAAAATGCTGAAGATGTTTGAAGCTTTCGAACCAA 1762
Qy      311 SERPROLEULEUGLUALARGARGLYANSPHEPHEULEUPHESERILEALAPRO 330
Db      1763 TCCCATGATGCGATGCGATGCGGCGGAGCAAGAACACTGACCTTATATGATGACCTT 1822
Qy      331 GLUVAL 332
Db      1823 GAGGTC 1828

```

## RESULT 8

```

US-08-923-454A-23
: Sequence 23. Application US/08923454A
: Patent No. 6004794
: GENERAL INFORMATION:
: APPLICANT: Creasy, Caretha
: APPLICANT: Liyi, George
: APPLICANT: Kairan, Eric
: APPLICANT: Clinkenbeard, Helen
: APPLICANT: Browne, Michael
: APPLICANT: Southan, Christopher
: TITLE OF INVENTION: HUMAN SERINE PROTEASE
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ Version 1.5
: CURRENT APPLICATION DATA:

```

```

: APPLICATION NUMBER: US/08/923,454A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/025436
: FILING DATE: 06-SEPT-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Baumeister, Kirk
: REGISTRATION NUMBER: 33,833
: REFERENCE/DOCKET NUMBER: P50547
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-5096
: TELEFAX: 610-270-5090
: TELEX:
: INFORMATION FOR SEQ ID NO: 23:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2187 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 603..1976
: OTHER INFORMATION:
: US-08-923-454A-23

```

## Alignment Scores:

```

Pred. No.: 5.92e-101 Length: 2187
Score: 882.00 Matches: 169
Percent Similarity: 74.84% Conservative: 72
Best Local Similarity: 52.48% Mismatches: 71
Query Match: 52.53% Indels: 10
Gaps: 2

```

US-10-617-443B-2 (1-334) x US-08-923-454A-23 (1-2187)

```

Qy      15 SERSERPROARGTYRGLYSAPHEASPHLEIALASPVVALGLUYSRIEALAPROALA 34
Db      1023 GCTTCTCCCGGAGCTGACGACATTTCATCGCAGATGTGTGAGAAAGACGACCTGCC 1082
Qy      35 VALVALHISILEGLIULEUPHEULARGHISPROLEUPHEGLYARGANVALPROLEUSER 54
Db      1083 GTGGCTATATCGAGATCTGACCGGACCCCTTCTGGGCGCGAGTCCCTATCTCG 1142
Qy      55 SERGLYSERGLYPHEILEMETSERGLUALGLYLEUILEIETHRANALAHISVALVAL 74
Db      1143 AACGGCTCAGAGATTCGGTGGCTGCGATGGGCTCATTTTCACCAACGCCCATGTGGTG 1202
Qy      75 SERSERANSERLAIALAPROGLYARGGLINLEULYSVALGLULEUGINANSGLYASP 94
Db      1203 GCTGAT-----CGGCGAGAGTCCGTGTGAGCTGCTAAGCGGCGAC 1244
Qy      95 SEPTYRGLUALATHRILEYASPILEYASPILEYSERAPITAEALATHRILEYILE 114
Db      1245 ACGATAGAGCCCGTGGTCAACGCTGTGATCCGCTGGACAGATCGCAACGCTGAGGATT 1304
Qy      115 HIEPROLYELYSLEUPROVALLEULEULEUGLYHISERIALASPLEUAPPROGLY 134
Db      1305 CAGACTAAGGAGCCTCTCCCAACGCTGCTGGAGCGCTCACTGATGTCCGGCAAGGG 1364
Qy      135 GUPHPEVALVALAIAIEGLISERPHHEALALEUGINAMTHVALTHTHGLYILE 154
Db      1365 GAGTTTGTGTGGCCATGGGAAGTCCCTTGGACGTGCAAGAACACATACATCCGGCATT 1424
Qy      155 VALSERTHRAGLAINARGLUGLYARGGLUBENGLYLEUARGPESERAPMETAPTFR 174
Db      1425 GTTAGCTGTGCTCAGGCTCCAGCAGACGACCTGGAGCTCCCAACCAATGTGAATAC 1484

```



Oy		175	IleGlnThrPrrApAlaIleIleAsnThyrGlyAnsSerGlyValProLeuValAsnLeuNsp	199
Dd		1485	ATTCAACTGTATGCAGCTATTTCATTGTTTGGAACTTGAGAGTCCCTCGTTAACCCTGGAT	1544
Oy		195	GlyGluValIleIleGlyIleAsnThrLeuLysValThrIlaGlyIleSerPheNlaIlePro	214
Dd		1545	GGGGAGGTGATGTGGATGAACACCATGAAAGGTCAACAGCTGGAAATCTCCTTGCCATCCCT	1604
Oy		215	SerAPArGIleThrArgPheLeuThrGluPheGlnAspLysGln-----Ile	230
Dd		1605	TCTGATGCTCTTCGAGAGTTCTGTGATGCTGGGGAAAAGAATCTCTCCTCGGAATC	1664
Oy		231	LysAPTrIpyLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeuVal	250
Dd		1665	AGTGAGTCCCAAGCGCGCTACATTTGGGGTAGATGTCGACCTGACGCCAGATCCTT	1724
Oy		251	AepGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyValGln	270
Dd		1725	GCTGAACCTACAGCTTCGAAACCAAGCTTTCGCCAATGTCAGCATGAGTCACTCATCCAT	1784
Oy		271	GluValAlaIaProAsnSerProSerGlnArgGlyIleGlnAspGlyASPRIleIleVal	290
Dd		1785	AAAGTCATCCCTGGGCTCCCCCTTCGACACCGGCTGTCTCGCGCTGGTGAITGATTTTG	1844
Oy		291	LysValAlaAngIYArgArgProLeuValAspSerSerGluLeuGlnGluValIleuThrGlu	310
Dd		1845	GCCATTGGGGAGCAGATGCTACAATAATCTGAAGATGTTTATGAAGCTGTTGCAACCCA	1904
Oy		311	SerProLeuLeuLeuGluValAlaArgGlyIlysAsnAspLeuLeuPheSerIleAlaPro	330
Dd		1905	TCCCAGTTGGCAGTGCAGATCCGGGGGAGACGAGAACACTGAACCTTATATATGACCCCT	1964
Oy		331	GluValI 332	
Dd		1965	GAGGTC 1970	
RESULT 9				
		US-08-923-454A-24		
		/ Sequence 24, Application US/08923454A		
		/ Patent No. 6004794		
		/ GENERAL INFORMATION:		
		/ APPLICANT: Creasy, Caretha		
		/ APPLICANT: Livi, George		
		/ APPLICANT: Kattan, Eric		
		/ APPLICANT: Clinkeneard, Helen		
		/ APPLICANT: Browne, Michael		
		/ APPLICANT: Southern, Christopher		
		/ TITLE OF INVENTION: HUMAN SERINE PROTEASE		
		/ NUMBER OF SEQUENCES: 40		
		/ CORRESPONDENCE ADDRESS:		
		/ ADDRESSEE: SmithKline Beecham Corporation		
		/ STREET: 709 Swedeland Road		
		/ CITY: King of Prussia		
		/ STATE: PA		
		/ COUNTRY: USA		
		/ ZIP: 19406		
		/ COMPUTER READABLE FORM:		
		/ MEDIUM TYPE: Diskette		
		/ COMPUTER: IBM Compatible		
		/ OPERATING SYSTEM: DOS		
		/ SOFTWARE: FastSeq Version 1.5		
		/ CURRENT APPLICATION DATA:		
		/ APPLICATION NUMBER: US/08/923,454A		
		/ FILING DATE:		
		/ CLASSIFICATION: 435		
		/ PRIOR APPLICATION DATA:		
		/ APPLICATION NUMBER: 60/025436		
		/ FILING DATE: 06-SEPT-1996		
		/ ATTORNEY/AGENT INFORMATION:		
		/ NAME: Baumeister, Kirk		
		/ REGISTRATION NUMBER: 33,833		
		/ REFERENCE/DOCKET NUMBER: P50547		
		/ TELECOMMUNICATION INFORMATION:		

```

: TELEPHONE: 610-270-5096
: TELERAX: 610-270-5090
: TELEX:
: INFORMATION FOR SEQ ID NO: 24:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2187 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 603...1976
: OTHER INFORMATION:
US-08-923-454A-24

Alignment Scores:
Pred. No.: 5.92e-101 Length: 2187
Score: 882.00 Matches: 169
Percent Similarity: 74.84% Conservative: 72
Best Local Similarity: 52.48% Mismatch: 71
Query Match: 52.53% Indels: 10
DB: 3 Gaps: 2

US-10-617-443B-2 (1-334) x US-08-923-454A-24 (1-2187)

Oy 15 SerSerProArgTgYLYvPheAenPhelEAlaaPvaValGULyVleAlaProLa 34
Db 1023 GCCTTCCTCCCGAGATCTGCACAACTTCACCAATGTGGTGGAAGAACAAGCAGCTGCC 1082
Oy 35 ValValHaillegLuleuPheLeuXhIaproleuPhegiLyagaenValProleuSer 54
Db 1083 GTGGCTATATCGAGATCTCTGACCAGCACCTTTCTTGGCCGCGAGAGTCCCTATCTCG 1142
Oy 55 SerGIserGIyPheilemetSerGIuaLagilyLeuileierlernaMaLiaVaivaI 74
Db 1143 AACGGCTCAGATTCTGATGTGGCTGCCAGATGGGCTCATTTGCAACAAAGCCCATGTGGTG 1202
Oy 75 SerSerenserialaalaProGIyArGdngInleuysvalGinleuginaSnglyAP 94
Db 1203 GCTGAT-----CGGCGAGATCGGTGTGAACATGCTMAAGCGGCGAC 1244
Oy 95 SerTyrgLuNaLThrileYAspIleaPylsySerAspIleaLThrileYle 114
Db 1245 ACGRHTAGGCCCGGTCACAGCTGTGATCCCTGGCAGACATCGAACCTGAGATT 1304
Oy 115 HieProlsylylvLeuProvalLeuLeuengIyhiSeralaaPleuXgProGIy 134
Db 1305 CAGACTAAGAGAGCTCTCCACCCGCTCTCTGGAGAGCTCAGCTGATGTCCGCAAGGG 1364
Oy 135 GluPheValaValaialiegIyserProPhalaLeuGlnaenThrValThrTrGIyle 154
Db 1365 GAGTTTGTGTTCGATCGAGAAAGTCCCTTTGCACTGCAGAACAGATCAATCCGGCATTT 1424
Oy 155 ValSerThraLaginaRgJugIyArGluDeuGIyLeuXrgasPserAspMetAsPyTyr 174
Db 1425 GTTAGCTCTGCTCAGCTCCAGCGACAGACCTGGGAGCTCCCCCAAACCAATGTGAATAC 1484
Oy 175 IlleGlnThrApriaalieleanttyGIyaSnserGIyGIProLeuValAnPleuAsp 194
Db 1485 ATTCAAACTAATGACGCTATTGATTTTGGAAACTGTGAAGTCTCCCTGTTAACCTGGAT 1544
Oy 195 GIyGUvalIllegIyleanThrLeuLyWaLThraLagilyleserPheAlalepro 214
Db 1545 GGCGAGGTGATTGGAGTGAACAACATAAGATGCACACTGGAATCTCTTTGCCATCCCT 1604
Oy 215 SerAspArgIleThraRphleuThnGluPheGlnaPylsgIn-----lle 230
Db 1605 TCTGATGTGCTTCGAGAGTTTCTGCATCTGTCGGGAAAAGAAATTCCTCTCCGGAATC 1664

```



QY 231 LysAspTrpLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeuVal 250  
 Db 1665 AGTGGGTCCTCCAGCGCGCTACATGGGTGATGATCTGACCTCCAGTCCAGCATCTT 1724  
 QY 251 AspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrValGln 270  
 Db 1725 GCTGAACACTACAGCTTCGAGAACCAAGCTTCCCGATGTTGACAGATGGTACTCAT 1784  
 QY 271 GluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleVal 290  
 Db 1785 AAAGTCATCTCGGCTCCCTGACACCGGGCTGCTGCGCGCTGGATGTATTTTG 1844  
 QY 291 LysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThrGlu 310  
 Db 1845 GCCATTGGGAGGACAGATGTCACAAATCTCAAGATGTTATGAACTGTTGCAACCA 1904  
 QY 311 SerProLeuLeuLeuGluValArgArgGlyLysAsnAspLeuLeuPheSerIleAlaPro 330  
 Db 1905 TCCAGATTGGCAGTGCAGATCCGGGGGAGCAGAAACACTGACCTTATGTGACCCCT 1964  
 QY 331 GluVal 332  
 Db 1965 GAGGTC 1970  
 RESULT 10  
 US-09-008-271A-23  
 Sequence 23, Application US/09008271A  
 Patent No. 6203979  
 GENERAL INFORMATION:  
 APPLICANT: Bandman, Olga  
 Hillman, Jennifer L.  
 Yue, Henry  
 Guegler, Karl J.  
 Corley, Neil C.  
 Tang, Tom Y.  
 Shah, Purvi  
 TITLE OF INVENTION: HUMAN PROTEASE MOLECULES  
 NUMBER OF SEQUENCES: 24  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Dr.  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/008.271A  
 FILING DATE: 16-Jan-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: <Unknown>  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mohan-Peterson, Sheela  
 REGISTRATION NUMBER: 41.201  
 REFERENCE/DOCKET NUMBER: PF-0458 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-855-0555  
 TELEFAX: 650-845-4166  
 INFORMATION FOR SEQ ID NO: 23:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2476 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: SINICT01  
 CLONE: 2680548

SEQUENCE DESCRIPTION: SEQ ID NO: 23 :  
 US-09-008-271A-23  
 Alignment Scores:  
 Pred. No.: 7.28e-101 Length: 2476  
 Score: 882.00 Matches: 169  
 Percent Similarity: 74.84% Conservative: 72  
 Best Local Similarity: 52.48% Mismatches: 71  
 Query Match: 52.53% Indels: 10  
 DB: 3 Gaps: 2  
 US-10-617-443B-2 (1-334) x US-09-008-271A-23 (1-2476)  
 QY 15 SerSerProArgTrpLysArgPheAsnPheIleAlaAspAlaValGluTyrIleAlaProAla 34  
 Db 481 GCTTCTCCCGAGCTCGTACAACTTATGCAATGATGTGTGGAGAAACAGACACCTGCC 540  
 QY 35 ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSer 54  
 Db 541 GTGGCTATATCGAGATCTCGAACCGGCACCTTTCTTG66CGCGAGTCCATCTCG 600  
 QY 55 SerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThrAsnAlaHisValVal 74  
 Db 601 AACGGCTCAGGATTCGTGTGCTGCGCATGGGCTCATTTGCACCAAGCCCATGTGATG 660  
 QY 75 SerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGlnLeuGlnAsnGlyAsp 94  
 Db 661 GCTGAT-----CGGCGAGAGTCCGTGTGACCTTAAGCGGCGAC 702  
 QY 95 SerTyrGluAlaThrIleLysAspIleAspLysSerAspIleAlaThrIleLysIle 114  
 Db 703 ACGTATAGGCGCGTGTGACAGCTGTGATCCCGTGCACACATCGCAACCTGAGATT 762  
 QY 115 HisProLysLysLysLeuProValLeuLeuGlnYHisSerAlaAspLeuArgProGly 134  
 Db 763 CAGACTAAGAGGCTCTCCCGACGCTGCTGGAGCGCTCAGCTGATGTCCGGCAAGGG 822  
 QY 135 GluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThrValIleThrGlyIle 154  
 Db 823 GAGTTGTGTTCGACGAGGAGAGCTCTTGTGCACTGAGAACACGATCAGATCCGGGATT 882  
 QY 155 ValSerThrAlaGlnArgGluGlyArgGluLeuGlyLysAspSerAspMetAspTyr 174  
 Db 883 GTTAGCTCTCTCAGCTCCAGCCAGAGACTGGAGCTCCCGCAACCATGTGGATTAC 942  
 QY 175 IleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyIleProLeuValAsnLeuAsp 194  
 Db 943 ATTCAAACTGATCAGCTATTGATTTGGAACTGTGAGAGTCCCTGTTAACCTGGAT 1002  
 QY 195 GlyGluValIleGlyIleAsnThrLeuLysValIleThrAlaGlyIleSerPheAlaIlePro 214  
 Db 1003 GGGAGGTGATTTGAGAGAACACCATGAGTACAGCTGGAATCTCTTTCATCTCC 1062  
 QY 215 SerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln-----Ile 230  
 Db 1063 TCTGATGCTCTTGCAGAGTTTCTGCATCTGCGGAAAGAAAGAAATCTCTCCGGAATC 1122  
 QY 231 LysAspTrpLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeuVal 250  
 Db 1123 AGTGGGTCCTCCAGCGCGCTACATGGGTGATGATGTCGACCTGATGCCACATCTT 1182  
 QY 251 AspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrValGln 270  
 Db 1183 GCTGAACACTACAGCTTCGAGAACCAAGCTTCCCGATGTTGACAGATGGTATCTCAT 1242  
 QY 271 GluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleVal 290  
 Db 1243 AAAGTCATCTCGGCTCCCTGACACCGGGCTGCTGCGCGCTGGATGTATTTTG 1302  
 QY 291 LysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThrGlu 310  
 Db 1303 GCCATTGGGAGGACAGATGTCACAAATGCTGAAGATGTTATGAACTGTTCAGAACCA 1362





APPLICANT: Creasey, Caretha  
APPLICANT: Livi, George  
APPLICANT: Karan, Eric  
APPLICANT: Clinkbeard, Helen  
APPLICANT: Browne, Michael  
APPLICANT: Southan, Christopher  
TITLE OF INVENTION: HUMAN SERINE PROTEASE  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/923,454A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/025436  
FILING DATE: 06-SEPT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumelster, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: P50547  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5096  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 603..1733  
OTHER INFORMATION:  
US-08-923-454A-26  
Alignment Scores:  
Pred. No.: 7,91e-95 Length: 2551  
Score: 834.50 Matches: 169  
Percent Similarity: 60.71% Conservative: 72  
Best Local Similarity: 42.57% Mismatches: 71  
Query Match: 49.70% Indels: 85  
Gaps: 3  
US-10-617-443B-2 (1-334) x US-08-923-454A-26 (1-2551)  
QY 15 SerSerProArgTyrIryshphenphenlealaaspvalaigluyllealeaProala 34  
DB 1023 GCTTCTCCCGGAGTACATCACTTCATCGAGTGTGTGAGAAACACACACTGCC 1082  
QY 35 ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgPheValProLeuSer 54  
DB 1083 GTGTCTATATCGAGATCCCGACCCGACCTTTCTTGCGCGCCAGGTCCCTATCTCG 1142  
QY 55 SerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThrAsnAlaHisValVal 74  
DB 1143 AACGGCTCAGAGATTCGTGTGCGCCGATGGCTCATGTGCACCAAGCCCATGTGTG 1202

QY 75 SerSerAsnSerAlaAlaProGlyArgGlnGlnLeuValGlnLeuGlnAsnGlyAsp 94  
DB 1203 GCTGAT-----CGGCGAAGATCCGTGTAGACTGTGAAGCGCGAC 1244  
QY 95 SerTyrGluAlaThrIleLeuAspIleAspLysSerAspIleAlaThrIleLysIle 114  
DB 1245 ACGTATGAGCCCGTGTACAGCTGTGATCCGCTGACAGATCGCAACGCTGAGATT 1304  
QY 115 HisProLysLysLysLeuProValLeuLeuGlyHisSerAlaAspLeuArgProGly 134  
DB 1305 CAGACTAAGGAGCTCTCCCGACGCTGCTGGAGCGCTCACTATGTCCGCGAAGG 1364  
QY 135 GluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThrValThrGlyIle 154  
DB 1365 GAGTTGTGTGTCCAGAGAAATCTTTGCACTGCAGAACAGATCAATCCGCGCAT 1424  
QY 155 ValSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyr 174  
DB 1425 GTTAGCTGTGCTCAGCGTCCAGAGACCTGGACTCCCGCAACCAATGTGAAATAC 1484  
QY 175 IleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeu 193  
DB 1485 ATCAAACTGATGCACTATGATTGGAATCTTGAGAGTCCCTGTTAACCCTGTG 1544  
QY 193 ----- 193  
DB 1545 AGTGAAGACATCTTCTCCAAAGATCCCTGCGCCAGAGTCACTGTGGAAAGGTAGTT 1604  
QY 193 ----- 193  
DB 1605 CCCCAATTCAAGATGTTGTGCAAGTTTGTAGCACTTCTTTGTTGGCTATCTCA 1664  
QY 193 ----- 193  
DB 1665 TATCCAAACCAATCTCCCAACACTGTGCTGTACTTTGTCGGGTGCCCCCATCCCTA 1724  
QY 194 -----AspGlyGluValIleGly 199  
DB 1725 CTATTGTTTACGCTAGGGAACGTGGGCGCTGATCTCCGACAGATGGGAGGTGATGGA 1784  
QY 200 IleAsnThrLeuLysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThr 219  
DB 1785 GTGAACACCATAGAGTCACTGAGATCTCTTGCATCCCTGATCGATGCTTCTGA 1844  
QY 220 ArgPheLeuThrGluPheGlnAspLysGln-----IleLysAspTrpLysLys 235  
DB 1845 GAGTTCTGCATCGTGGGGAAGAAATCTCTCCCGAATCATGGGTCCACGCG 1904  
QY 236 ArgPheIleGlyIleArgMetArgThrIleThrProSerLeuValAspGluLeuValAla 255  
DB 1905 CGCTAATTGGGGGTGATGCTGACCTCGAATCCAGATCTTCGTAACCTACAGCTT 1964  
QY 256 SerAsnProAspPheProGluValSerSerGlyIleTyrValGlnGluValAlaProAsn 275  
DB 1965 CGAAGAACCAAGCTTCCCATGTTTCAGATGTTCACTCATCAATCAATCAATCTCGGC 2024  
QY 276 SerProSerGlnArgGlyIleGlnAspGlyAspIleIleValLysValAsnGlyArg 295  
DB 2025 TCCCTGACACACCGGCGCTGCTGCGCGCTGATGATGATTTTGCCCATTTGGGAGCAG 2084  
QY 296 ProLeuValAspSerSerGluLeuGlnGluValAlaLeuThrGluSerProLeuLeuLeu 315  
DB 2085 ATGTATCAAAATCTGAAAGATGTTTGAAGCTGTTCGAACCCAAATCCAGTTGGCAGTG 2144  
QY 316 GluValArgArgLysAsnAspLeuPheSerIleAlaProGluVal 332  
DB 2145 CAGATCCGGCGGAGCGAAGAACATGACTTATATGTACCCCTTAGGTTC 2195  
RESULT 14  
US-08-923-454A-5  
Sequence 5, Application US/08923454A  
Patent No. 6004794



```

ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,454A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/025436
FILING DATE: 06-SEPT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50547
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 2144 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 603...1910
OTHER INFORMATION:
US-08-923-454A-28

```

```

Alignment Scores:
Pred. No.: 6,33e-88 Length: 2144
Score: 779.00 Matches: 158
Percent Similarity: 66.87% Conservative: 64
Best Local Similarity: 47.59% Mismatches: 58
Query Match: 46.40% Indels: 52
DB: 3 Gaps: 4

```

US-10-617-443B-2 (1-334) x US-08-923-454A-28 (1-2144)

```

QY 15 SerSerProAArgTyrLysPheAsnPhelLeAlaAPValValGluLysIleAlaProAla 34
DB 1023 GCTTCTCCCGAGTCAGTACACTTCAATCGAGATGTGTGGAGAGACACGACCTGCC 1082
QY 35 ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSer 54
DB 1083 GTGGCTATATGAGATCTCGACCGGACCTTTCTGGGCGCGGAGTCTCTATCTCG 1142
QY 55 SerGlySerGlyPheIleMetSerGluValaglyLeuIleIleThrAsnAlaHisValAl 74
DB 1143 AACGGCTCAGATTCGTGTGTGTCGCCGATGGCTCATTTGTCACCAACGCCCATGTGTG 1202
QY 75 SerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGlnLeuGlnAsnGlyAsp 94
DB 1203 GCTGAT-----CGGCGCAGATCCGTGTGAGCTGCTAAGCGCGGAC 1244
QY 95 SerTyrGluAlaThrIleLysAspIleAspLysSerAspIleAlaThrIleLysIle 114
DB 1245 ACGTATGAGCGCGGTGTCAAGCTGTGATCCCGTGGCAGACATCGCAACGCTGAGATT 1304
QY 115 HisProLysLysLeuLeuProValLeuLeuLysGlyHisSerAlaAspLeuArgProGly 134
DB 1305 CAGACTTAAGAGCTCTCCCGACGCTGTGGAGCGCTCAGCTCATCTCCGCGAAGGG 1364
QY 135 GluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThrValThrThrGlyIle 154

```

```

DB 1365 GAGTTGTGTGTCATGAGAGTCCCTTTCATCTGCACAGACATCATCCGCGCAT 1424
QY 155 ValSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyr 174
DB 1425 GTTAGCTCTGCTCAGGCTCAGCCAGACCTGGGACTCCCCCAACCATATGTGAAATAC 1484
QY 175 IleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeu--- 193
DB 1485 ATTCAACTGATCAGCTATGATTGTAATTTGGAACTCTGAGAGTCCCTGTGTTAACCTGGCT 1544
QY 194 -----AspGlyGluValIleGlyIleAsnThrLeuLys 204
DB 1545 AGGAACTGGGGGCTGATCCCTGCAGAGATGGGAGATGATGAGTGAACACCATGAAG 1604
QY 205 ValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThrGlu 224
DB 1605 GTCACTGAGATCTCTTGGCATCCCTTGATGCTCTTCAGAGATTTCGTCATCGT 1664
QY 225 PheGlnAspLysGln-----IleLysAspTyrLysArgPheIleGlyIle 240
DB 1665 GGGAAAGAAAGAAATTCTCTCCGGAATCAGTGGGTCCGAGCGCGCTCATTTGGGGTG 1724
QY 241 ArgMetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAsnProAspPhe 260
DB 1725 ATGATGCTGACCTGAGTCC----- 1745
QY 261 ProGluValSerSerGlyIleTyrValGlnGluValAlaProAsnSerProSerGlnArg 280
DB 1746 -----AGC 1748
QY 281 GlyGlyIleGlnAspGlyAspIleIleValLysValAsnGlyArgProLeuValAspSer 300
DB 1749 GCTGCTCGCGCGCTGGTGTGATTTTGGCCATTGGGAGCAGATGATCAAAATGCT 1808
QY 301 SerGluLeuGlnGluAlaValLeuThrGluSerProLeuLeuLeuGluValArgArgGly 320
DB 1809 GAAGATGTTTGAAGCTGTTTCGAACCCCAATCCAGATGGCAGATCCGCGGGGA 1868
QY 321 AsnAspAspLeuLeuPheSerIleAlaProGluVal 332
DB 1869 CGAGAAACATGACTTATATGTGACCCCTGAGGTC 1904

```

Search completed: June 30, 2005, 01:32:59  
Job time : 320.971 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_p1ue\_p2n model

Run on: June 29, 2005, 21:23:47; Search time 1045.82 Seconds

(without alignments)  
1996.208 Million cell updates/sec

Title: US-10-617-443B-2

Perfect score: 1679  
Sequence: 1 MHIALPASAGLHQUSPRYK.....LEVRRGNDLLFSIAPEVVM 334

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 6067389 seqs, 3125258755 residues

Total number of hits satisfying chosen parameters: 12134778

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=x1h  
-Q=/cgn2\_1/USPTO.epool/US10617443/runat\_29062005\_135933\_26126/app\_query.fasta\_1.718  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=trmp -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blcsum62  
-TRANS=human40.csl -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10617443@cgn\_1\_1\_582@runat\_29062005\_135933\_26126  
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLCCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database:

Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/1/pubpna/US10J\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/1/pubpna/US10K\_PUBCOMB.seq:\*  
23: /cgn2\_6/ptodata/1/pubpna/US10L\_PUBCOMB.seq:\*  
24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1679	100.0	3006	17	US-10-189-099A-1
2	1679	100.0	3006	21	US-10-617-443B-1
3	1635	97.4	1044	9	US-09-796-858-9
4	1635	97.4	1563	10	US-09-796-753-3
5	1635	97.4	1797	18	US-10-275-505-28
6	1635	97.4	2040	9	US-09-818-143-14
7	1635	97.4	2094	19	US-10-381-820A-7
8	1635	97.4	2541	21	US-10-356-157-5181
9	1635	97.4	2543	21	US-10-485-313A-31
10	1635	97.4	2576	10	US-09-796-753-31
11	1536	91.5	2450	21	US-10-485-313A-26
12	1482.5	88.3	2551	15	US-10-301-822-78
13	1378	82.1	1868	9	US-09-764-898-112
14	1122	66.8	1695	9	US-09-969-384-7
15	1117	66.5	1953	21	US-10-485-313A-32
16	1105	65.8	1788	9	US-09-764-898-40
17	1057	63.0	1897	21	US-10-485-313A-38
18	1034.5	61.6	1440	9	US-09-796-858-7
19	1034.5	61.6	1894	17	US-10-104-047-795
20	1034.5	61.6	2034	15	US-10-102-524-1732
21	1034.5	61.6	2036	10	US-09-960-706-516
22	1034.5	61.6	2036	17	US-10-170-385-356
23	1034.5	61.6	2036	17	US-10-172-118-815
24	1034.5	61.6	2036	18	US-10-342-887-815
25	1034.5	61.6	2036	21	US-10-772-636-47
26	1034.5	61.6	2036	21	US-10-956-157-919
27	1034.5	61.6	2054	15	US-10-084-817-32
28	1034.5	61.6	2437	14	US-10-198-846-10826
29	1027.5	61.2	2157	17	US-10-264-049-135
30	1006.5	59.9	1855	9	US-09-925-298-134
31	1006.5	59.9	1855	14	US-10-102-806-134
32	1005.5	59.9	2212	20	US-10-723-860-5299
33	981.5	58.5	1534	9	US-09-765-231A-20
34	945	56.3	1388	21	US-10-764-420-1506
35	920	54.8	1859	14	US-09-866-050A-588
36	920	54.8	1859	14	US-10-152-661-588
37	890	53.0	1544	19	US-10-789-241-49
38	890	53.0	1544	21	US-10-772-636-7
39	890	53.0	1553	9	US-09-935-390A-18
40	882	52.5	963	19	US-10-730-476A-11
41	882	52.5	975	19	US-10-730-476A-1
42	882	52.5	1374	14	US-10-197-63A-14
43	882	52.5	1400	21	US-10-956-157-7054
44	882	52.5	1597	17	US-10-352-684A-9
45	882	52.5	1597	21	US-10-956-157-1819

## ALIGNMENTS

RESULT 1  
US-10-189-099A-1  
; Sequence 1, Application US/10189099A  
; Publication No. US20040005659A1  
; GENERAL INFORMATION:  
; APPLICANT: Darrow, Andrew L  
; APPLICANT: Qi, Jian-shen  
; APPLICANT: Chen, Caillin  
; APPLICANT: Andrade-Gordon, Patricia  
; TITLE OR INVENTION: Human PRSS-11 like S2 serine protease and uses thereof  
; FILE REFERENCE: ORT-1644  
; CURRENT APPLICATION NUMBER: US/10/189, 099A  
; CURRENT FILING DATE: 2002-12-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1



```

; LENGTH: 3006
; TYPE: CDNA
; ORGANISM: Homo sapiens
US-10-189-099A-1

Alignment Scores:
Pred. No.: 2,43e-180 Length: 3006
Score: 1679.00 Matches: 334
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-617-443B-2 (1-334) x US-10-189-099A-1 (1-3006)

QY 1 MethiLeuAlaLeuProAlaSerAlaGlyLeuHiGlnLeuSerSerProArgTyrLys 20
DB 1011 ATGCACCTGACCTTCCCGCCAGCGAGCTCTCCACAGCTGAGCCCGCTCAAG 1070
QY pheAenPheIleAlaAspValGluValIleAlaProAlaValAlaHisIleGluLeu 40
DB 1071 TTCACCTTATGCTGACGTTGGAGAAAGATGCACCGCTGATCCACATAGAGCTC 1130
QY pheLeuArgHisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIle 60
DB 1131 TTCCTGAGACACCCCGCTTGGCCGCAAGTGCCTGTCAGCGGCTTCTGCTTCATC 1190
QY MetSerGlnAlaGlyLeuIleIleThrAsnAlaHisValIleSerSerAsnSerAlaIle 80
DB 1191 ATGTGAGAGCCGCTGATCATCACCAATGCCAGTGTGTCAGCAACAGTGTGCC 1250
QY ProGlyArgGlnGlnLeuLysValGlnLeuGlnAsnGlyAspSerTyrGlnAlaThrIle 100
DB 1251 CCGGGGAGGACAGCTCAAGGTGCAGCTACAGAACTGGGACTCTCTTAAGCCCAATC 1310
QY LysAspIleAspIleLysSerAspIleAlaThrIleLysIleHisProLysIleLysLeu 120
DB 1311 AAAGACATGACACAAAGATCGGACATTCGCACCATCAATCCATCCCAAGAAAAGCTC 1370
QY ProValIleuLeuLeuGlyHisSerAlaAspLeuArgProGlyGluPheValAlaIle 140
DB 1371 CCGTGTGTTTGTCTGGGTCACTCGGCCGACCTCGGCGGAGTTTGTGGGCCATC 1430
QY GlySerProPheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArg 160
DB 1431 GGCAAGCTCTTGGCTTACAGAAACAGACAGCAAGCGGATGTCAGCATGCGCCAGCGG 1490
QY GluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIle 180
DB 1491 GAGGAGGAGGAGCTGGGCTCCGCGGACTCCGACATGACATACATCCAGACGATGCCATC 1550
QY IleAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleGlyIle 200
DB 1551 ATCACTACGGGAACTCCGGGGACCACTGTGAACCTGATGGCCAGGATCATTTGGCATC 1610
QY AsnThrLeuLysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArg 220
DB 1611 AACACGCTAAGGTCAAGCTGGCATCTCTTTGCCATCTCCCTCAGACCGCATCACAGG 1670
QY PheLeuThrGluPheGlnAspIleLysGlnIleLysAspTyrLysIleValArgPheIleGlyIle 240
DB 1671 TTCCTCAGACAGATTCCAAAGACAAAGCATCAAGATCGAAGAGAGCGTTTCATCGGCATA 1730
QY ArgMetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAsnProAspPhe 260
DB 1731 CGGATGCGGACGATCACCAAGCTCGTGGATGAGCTAAGCCGCAACCCGCGACTTC 1790
QY ProGluValSerSerGlyIleTyrValGlnGluValAlaProAsnSerProSerGlnArg 280
DB 1791 CCGAGGTGACGAGTGGATTTATGTGCAAGAGTTGCGCCGAAATTCACCTTCTCAGAGA 1850
QY GlyGlyIleGlnAspGlyAspIleIleValIleValAsnGlyArgProLeuValAspSer 300
DB 281 GlyGlyIleGlnAspGlyAspIleIleValIleValAsnGlyArgProLeuValAspSer 300
```

```

DB 1851 GGCGGATCAACAGATGTGATCATCTCAAGGTCAACGGGCGTCTCTAGGACTCG 1910
QY SerGluLeuGlnGluAlaValLeuThrGluSerProLeuLeuGluValArgArgGly 320
DB 1911 AGTGAGCTGACGAGAGCGGTGCTGACCGAGTCTCTCTTCACTGAGAGGTGGCGGAGG 1970
QY 321 AsnAspAspLeuLeuPheSerIleAlaProGluValIleMet 334
DB 1971 AACGACGACCTCTCTTCAAGCATCGCACCTGAGGTGATCATG 2012

RESULT 2
US-10-617-443B-1
; Sequence 1, Application US/10617443B
; Publication No. US20050019777A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-Shen
; APPLICANT: Chen, Cailin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: CRT-1644
; CURRENT APPLICATION NUMBER: US/10/617,443B
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3006
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-617-443B-1

Alignment Scores:
Pred. No.: 2,43e-180 Length: 3006
Score: 1679.00 Matches: 334
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-10-617-443B-2 (1-334) x US-10-617-443B-1 (1-3006)

QY 1 MethiLeuAlaLeuProAlaSerAlaGlyLeuHiGlnLeuSerSerProArgTyrLys 20
DB 1011 ATGCACCTGACCTTCCCGCCAGCGAGCTCTCCACAGCTGAGCCCGCTCAAG 1070
QY pheAenPheIleAlaAspValGluValIleAlaProAlaValAlaHisIleGluLeu 40
DB 1071 TTCACCTTATGCTGACGTTGGAGAAAGATGCACCGCTGATCCACATAGAGCTC 1130
QY pheLeuArgHisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIle 60
DB 1131 TTCCTGAGACACCCCGCTTGGCCGCAAGTGCCTGTCAGCGGCTTCTGCTTCATC 1190
QY MetSerGlnAlaGlyLeuIleIleThrAsnAlaHisValIleSerSerAsnSerAlaIle 80
DB 1191 ATGTGAGAGCCGCTGATCATCACCAATGCCAGTGTGTCAGCAACAGTGTGCC 1250
QY ProGlyArgGlnGlnLeuLysValGlnLeuGlnAsnGlyAspSerTyrGlnAlaThrIle 100
DB 1251 CCGGGGAGGACAGCTCAAGGTGCAGCTACAGAACTGGGACTCTCTTAAGGCCCAACATC 1310
QY LysAspIleAspIleLysSerAspIleAlaThrIleLysIleHisProLysIleLysLeu 120
DB 1311 AAAGACATGACACAAAGATCGGACATTCGCACCATCAATCCCAAGAAAAGCTC 1370
QY GlySerProPheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArg 160
DB 1431 GGCAAGCTCTTGGCTTACAGAAACAGACAGCAAGCGGATGTCAGCATGCGCCAGCGG 1490
```



Qy	161	GIUGLYA <sup>RG</sup> LU <sup>LEUG</sup> LU <sup>LEU</sup> ARG <sup>SPSE</sup> ASP <sup>ME</sup> LAP <sup>RY</sup> LIE <sup>GI</sup> TH <sup>ASP</sup> ALL <sup>LE</sup>	180
Db	1491	GAGGCGAAGGACTGGGCTCCGGGACTCCGACATGACTACATCAACGGATGCCATC	1550
Qy	181	IL <sup>AS</sup> NT <sup>RY</sup> GI <sup>YAS</sup> N <sup>SE</sup> RG <sup>LY</sup> PRO <sup>LEU</sup> VAL <sup>AE</sup> LU <sup>AE</sup> ASP <sup>GL</sup> YGL <sup>VAL</sup> IL <sup>EG</sup> LY <sup>LE</sup>	200
Db	1551	ATCAACTACGGGAACTCCGGGGGACCATCGTGAACTCGATGGGAGGTCAATGGCATC	1610
Qy	201	AE <sup>NT</sup> HR <sup>LEU</sup> VAL <sup>TH</sup> AL <sup>GL</sup> Y <sup>LE</sup> SE <sup>R</sup> PH <sup>EA</sup> L <sup>IL</sup> EP <sup>RO</sup> SE <sup>R</sup> ASP <sup>ARG</sup> IL <sup>ETH</sup> ARG	220
Db	1611	AACAGCTCAAGGTACAGGCTGGCATCTCTTGGCATCCCTCAGACCGCATCAACGG	1670
Qy	221	PH <sup>EU</sup> NT <sup>HR</sup> GI <sup>UP</sup> HE <sup>GL</sup> NA <sup>SP</sup> LY <sup>AG</sup> NI <sup>LE</sup> YS <sup>AP</sup> TR <sup>PL</sup> Y <sup>LY</sup> AS <sup>Y</sup> RP <sup>HE</sup> IL <sup>EG</sup> LY <sup>LE</sup>	240
Db	1671	TYCTCTCAAGAGTTCCCAAGACACGACGATCAAGCTGAAGAGAACGGCTTCATCGGCATA	1730
Qy	241	ARG <sup>ME</sup> T <sup>ARG</sup> TR <sup>IL</sup> ETH <sup>R</sup> PRO <sup>SE</sup> RL <sup>EU</sup> VAL <sup>ASP</sup> GI <sup>LU</sup> VAL <sup>YAL</sup> AS <sup>ER</sup> ASP <sup>AP</sup> HE	260
Db	1731	CGGATGCCGACGATCACACCAAGCTGTGGATGAGCTGAAGGCGCAAGCAACCGGACATTC	1790
Qy	261	PRO <sup>GI</sup> LU <sup>AL</sup> SER <sup>SE</sup> RG <sup>LY</sup> LE <sup>TY</sup> RV <sup>AL</sup> GI <sup>NG</sup> LU <sup>VAL</sup> AI <sup>PRO</sup> AS <sup>ER</sup> PRO <sup>SE</sup> RG <sup>LI</sup> N <sup>ARG</sup>	280
Db	1791	CCAAGGCTCACACAGTGAATTAATGTGCAAGAGGTGGCCGAATTCATCTTCAAGAGA	1850
Qy	281	GLY <sup>GL</sup> Y <sup>IL</sup> EG <sup>LI</sup> NA <sup>SP</sup> LY <sup>ASP</sup> IL <sup>EL</sup> LE <sup>VAL</sup> Y <sup>VAL</sup> AE <sup>NG</sup> LY <sup>ARG</sup> PRO <sup>LEU</sup> VAL <sup>ASP</sup> SE <sup>R</sup>	300
Db	1851	GGCGGCATCCAAAGATGTGCATCATCTGCAAGGTCAACGGGCGCTCTAGTGACATCG	1910
Qy	301	SE <sup>RI</sup> LU <sup>ENG</sup> ING <sup>LU</sup> AL <sup>VAL</sup> LEU <sup>TH</sup> RG <sup>LI</sup> SE <sup>R</sup> PRO <sup>LEU</sup> LEU <sup>ENG</sup> LU <sup>VAL</sup> ARG <sup>ARG</sup> LY	320
Db	1911	AGTAGCTGCAGAGGCGGTGCTGACCAAGTCTCTCTCCCTCACTGAAGAGTGGCGGGGG	1970
Qy	321	ASN <sup>ASP</sup> AS <sup>PL</sup> LEU <sup>AE</sup> PH <sup>SE</sup> RI <sup>L</sup> AL <sup>AR</sup> PRO <sup>GI</sup> LU <sup>VAL</sup> IM <sup>ET</sup>	334
Db	1971	AACGACGACCTCTCTTTCAGCATGCGACCTTAGGTGGTCAAG	2012

```

RESULT 3
US-09-796-858-9
Sequence 9, Application US/09796858
Patent No. US20020055139A1
GENERAL INFORMATION:
APPLICANT: Holtzmann, Douglas
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING PROGNOSTIC, DIAGNOSTIC,
TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
FILE REFERENCE: 7853-226-999
CURRENT APPLICATION NUMBER: US/09/796,858
CURRENT FILING DATE: 2001-03-01
PRIORITY APPLICATION NUMBER: 09/223,094
PRIORITY FILING DATE: 1998-12-30
PRIORITY APPLICATION NUMBER: 09/223,546
PRIORITY FILING DATE: 1998-12-30
PRIORITY APPLICATION NUMBER: 09/224,246
PRIORITY FILING DATE: 1998-12-30
PRIORITY APPLICATION NUMBER: 09/312,359
PRIORITY FILING DATE: 1999-05-14
PRIORITY APPLICATION NUMBER: 09/336,536
PRIORITY FILING DATE: 1999-06-18
PRIORITY APPLICATION NUMBER: 09/342,687
PRIORITY FILING DATE: 1999-06-29
PRIORITY APPLICATION NUMBER: 09/399,723
PRIORITY FILING DATE: 1999-09-20
PRIORITY APPLICATION NUMBER: 09/471,179
PRIORITY FILING DATE: 1999-12-23
PRIORITY APPLICATION NUMBER: 09/474,071
PRIORITY FILING DATE: 1999-12-29
PRIORITY APPLICATION NUMBER: 09/474,072
PRIORITY FILING DATE: 1999-12-29
PRIORITY APPLICATION NUMBER: 09/572,002
PRIORITY FILING DATE: 2000-05-14
PRIORITY APPLICATION NUMBER: 09/597,993
PRIORITY FILING DATE: 2000-06-12
PRIORITY APPLICATION NUMBER: 09/599,596
PRIORITY FILING DATE: 2000-06-22

```

```

? PRIOR APPLICATION NUMBER: 09/606,565
? PRIOR FILING DATE: 2000-06-29
? PRIOR APPLICATION NUMBER: 09/365,164
? PRIOR FILING DATE: 1998-07-30
? PRIOR APPLICATION NUMBER: 09/630,334
? PRIOR FILING DATE: 2000-07-31
? PRIOR APPLICATION NUMBER: 09/665,666
? PRIOR FILING DATE: 2000-09-20
? NUMBER OF SEQ ID NOS: 50
? SEQ ID NO 9
? LENGTH: 1044
? TYPE: DNA
? ORGANISM: Mouse
US-09-796-858-9

Alignment Scores:
Pred. No.: 5,42e-176
Score: 1635.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 97.38%
DB: 9

```

```

.Length: 1044
Matches: 325
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

```

US-10-617-443B-2 (1-334) X US-09-796-858-9 (1-1044)

Qy	10	GI	Vleu	HI	GI	Leu	Sen	Se	Pro	Arg	Trp	Leu	Phen	Ala	Leu	Val	GI	29
Db	70	GG	CT	CC	AC	CA	CG	TA	GC	AG	CC	CG	CT	CA	AG	TT	CA	1239
Qy	30	LY	SI	LA	AP	RO	LA	VA	LA	HI	SI	LE	GU	LEU	PH	EU	ARG	49
Db	130	AA	GA	TG	CA	CA	CG	CG	TG	GT	CC	CA	TAG	AG	CT	TT	CG	189
Qy	50	AA	SN	VA	L	P	RO	LEU	SEN	SE	RG	LY	SE	RG	LY	PH	ET	69
Db	190	AA	AG	T	C	C	C	C	T	G	C	A	G	C	G	G	T	249
Qy	70	AA	SI	LA	HI	SI	VA	LA	SE	RG	SEN	SE	LA	LA	P	RO	GI	89
Db	250	AA	TG	CC	CA	CG	TG	GT	CT	CA	CA	AG	CT	GC	CC	CG	CA	309
Qy	90	LEU	GI	AA	NG	I	AA	SP	SE	TR	GI	LA	HI	LEU	SE	AP	LE	109
Db	310	CT	AC	AA	AA	I	GG	GA	CT	CT	CT	TA	AG	GC	CA	CA	735	
Qy	110	LA	TH	RI	LEU	SI	LE	HI	SE	RG	LY	SE	LA	LEU	LEU	SE	129	
Db	370	GC	CA	CC	AT	CA	AA	GC	AT	CC	CA	AA	AA	AG	CT	CC	429	
Qy	130	AA	SP	LEU	AA	P	RO	GI	LEU	VA	LA	HI	LEU	SE	RG	LY	149	
Db	430	GA	CT	CG	CG	CT	CG	GA	AG	TT	GT	GT	GC	CA	CT	CG	489	
Qy	150	VA	L	TH	TR	GI	L	VA	L	SE	RG	LA	NG	I	AA	RG	169	
Db	490	GT	TA	CA	CA	CG	GA	TG	TA	CA	CG	CC	AG	GG	CA	GG	549	
Qy	170	SE	RG	ME	CA	SP	TR	GI	LEU	SE	LA	LEU	SE	RG	LY	PRO	189	
Db	550	TC	GA	CA	TG	GA	CT	CA	CA	CG	GA	TG	CA	CA	CT	CG	609	
Qy	190	LEU	VA	L	AA	SN	VA	L	GI	LEU	PH	EU	ARG	TR	LEU	SI	209	
Db	610	CT	GT	TA	AA	CT	GT	GA	TG	CG	AG	CT	TA	CA	CA	CG	669	
Qy	210	SE	R	PH	LA	L	E	P	R	O	S	E	R	G	L	E	229	
Db	670	TC	CT	T	T	G	C	A	T	C	C	C	T	CA	CA	CG	729	
Qy	230	L	E	Y	A	S	P	T	R	Y	L	E	Y	A	R	G	249	
Db	730	AT	CA	AA	GA	CT	GA	AA	AG	CT	CT	CA	TG	CA	TG	CA	789	



Qy 250 ValaapgluleuylalaSeranProaspheProgluValaSerSerGlyIleTyVal 269  
Db GTGATGAGCTGAGAGCCAGCAACCGACTTCCAGAGGTGACAGAGTGAATTATATG 1116  
Qy 270 GlnGluValaIaProanSerProserGlnArgGlyIleGlnaPglYaSpIleIle 289  
Db CAAGAGGTGGCCCAATTCACCTTCAGAGAGCGGCATCCAAAGATGTGATCATATC 1176  
Qy 290 ValaValaIaPglYaSpProleuValaPheSerSerGluGlnGlnaIaValaIleThr 309  
Db GTCAAGGTCAACGGCGCTCTTACTGAGCTGAGTGAAGTGAAGAGCGGTGCTGACC 1236  
Qy 310 GluSerProleuLeuGluValaIaArgGlyAaenAaPleuPheSerIleAla 329  
Db GAGTCTCTCTCTACTGAGAGTGGCGGGGAAAGCAAGACTCTTCAAGCATGCCA 1296  
Qy 330 ProgluValaIaMet 334  
Db CCTGAGGTGTCATG 1311

## RESULT 5

US-10-275-505-28  
; Sequence 28, Application US/10275505  
; Publication No. US20040081961A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: DEBEGRANT, Angelo M.; LAL, Preeti G.  
; APPLICANT: HAFALIA, April J.A.; PATTERSON, Chandra  
; APPLICANT: WALIA, Nardinder K.; KEARNEY, Liam  
; APPLICANT: TRIBOULEY, Catherine M.; KHAN, Farrah A.  
; APPLICANT: YAO, Monique G.; BAUGHN, Mariah R.  
; APPLICANT: AZIMZAI, Yalda; ELLIOTT, Vicki S.  
; APPLICANT: NGUYEN, Daniel B.; GANDHI, Ameena R.  
; APPLICANT: YANG, Junning; HERNANDEZ, Roberto  
; APPLICANT: POLICKY, Jennifer L.; LU, Dying Aina M.  
; APPLICANT: REDDY, Roopa M.; YUE, Henry  
; APPLICANT: TANG, Y. Tom  
; TITLE OF INVENTION: PROTEASES  
; FILE REFERENCE: PI-0085 USN  
; CURRENT APPLICATION NUMBER: US/10/275, 505  
; PRIOR FILING DATE: 2002-11-04  
; PRIOR APPLICATION NUMBER: PCT/US01/14651  
; PRIOR FILING DATE: 2001-05-04  
; PRIOR APPLICATION NUMBER: 60/209,402  
; PRIOR FILING DATE: 2000-06-01  
; PRIOR APPLICATION NUMBER: 60/207,477  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: 60/205,803  
; PRIOR FILING DATE: 2000-05-17  
; PRIOR APPLICATION NUMBER: 60/203,566  
; PRIOR FILING DATE: 2000-05-11  
; PRIOR APPLICATION NUMBER: 60/202,082  
; PRIOR FILING DATE: 2000-05-04  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: PERL Program  
; SEQ ID NO 28  
; LENGTH: 1797  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No: 7474343CBI  
US-10-275-505-28

## Alignment Scores:

Pred. No.: 1.19e-175 Length: 1797  
Score: 1635.00 Matches: 325  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 97.38% Indels: 0  
Gaps: 0  
18

US-10-617-443B-2 (1-334) x US-10-275-505-28 (1-1797)

Qy 10 GlyLeuHleGlnLeuSerSerProArgTyrlaPhePhePheIleIaAaPglValaGlu 29  
Db GGCTCCACACCGAGTACAGAGCCCGCTCAACAAGTTCAATTCATTCGTGAGCTGAGTGGAG 624  
Qy 30 LysIleAaProaIaValaIaIleIleGluPheLeuArgHleProLeuPheGlyArg 49  
Db AAGATCCACACAGCCGCTGCTCAATAGAGCTCTTCTAGACACCCGCTGTTGGCCG 684  
Qy 50 AaValProleuSerSerGlySerGlyPheIleLeuSerGluAaGlyLeuIleThr 69  
Db AACGTCCCTCTGTCACAGGCTTGCTTCATCATGTCAAGAGCGCGCTCATCATCAACC 744  
Qy 70 AaValaIaValaIaSerSerAaSerAlaAaProIaArgGlnGlnLeuValaGln 89  
Db AATGCCACAGGTGTCTCAGACACAGTGTGCCCCGGGAGGAGGAGCTCAAGGTGACG 804  
Qy 90 LeuGlnAaGlyAaPleuSerTyrgluAlaThrIleAaPleuAaPleuAaPleuAaPleu 109  
Db CTACAGAAATGGAGCTCTTATGAGCCACCATCAAAAGCATGACAAAGAGTCCGACATT 864  
Qy 110 AlaThrIleAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleu 129  
Db GCCACCATCAAGATCCATCCCAAGAAAAGCTCCCTGTGTGTGTGTGTGTGTGTGTGTGT 924  
Qy 130 AaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleu 149  
Db GACCTGGGGCTGGGGAGTTGT 984  
Qy 150 ValaThrGlyIleValaSerThrAlaGlnArgGluGlnArgGluGlnArgGluGlnArg 169  
Db GTGCAACGGGCATGCTGACGACCTGCCAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1044  
Qy 170 SerAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleu 189  
Db TCCGACATGAGCTACATCAAGAGGATCCATCATCAATCAAGGAACTCCGGGGAGCA 1104  
Qy 190 LeuValaAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleu 209  
Db CTGTGTAACCTGATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1164  
Qy 210 SerPheAaIaIleProSerAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleu 229  
Db TCCCTTCCATCCCTCTAGACCGCATACACGTTCTCTCAAGATTCCAAAGACACAG 1224  
Qy 230 IleAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleuAaPleu 249  
Db ATCAAAAGCTGGAAGAGGCGCTTATGCGCATACGATGCGAGATGACACCAAGCGCTG 1284  
Qy 250 ValaapgluleuylalaSeranProaspheProgluValaSerSerGlyIleTyVal 269  
Db GTGATGAGCTGAGAGCCAGCAACCGACTTCCAGAGGTGACAGAGTGAATTATATG 1344  
Qy 270 GlnGluValaIaProanSerProserGlnArgGlyIleGlnaPglYaSpIleIle 289  
Db CAAGAGGTGGCCCAATTCACCTTCAGAGAGCGGCATCCAAAGATGTGATCATATC 1404  
Qy 290 ValaValaIaPglYaSpProleuValaPheSerSerGluGlnGlnaIaValaIleThr 309  
Db GTCAAGGTCAACGGCGCTCTTACTGAGCTGAGTGAAGTGAAGAGCGGTGCTGACC 1464  
Qy 310 GluSerProleuLeuGluValaIaArgGlyAaenAaPleuPheSerIleAla 329  
Db GAGTCTCTCTCTACTGAGAGTGGCGGGGAAAGCAAGACTCTTCAAGCATGCCA 1524  
Qy 330 ProgluValaIaMet 334  
Db CCTGAGGTGTCATG 1539

## RESULT 6

US-09-818-143-14  
; Sequence 14, Application US/09818143

Patent No. US20020019000A1  
 GENERAL INFORMATION:  
 APPLICANT: Walker, Michael G.  
 APPLICANT: Volkmut, Wayne  
 APPLICANT: Klingler, Tod M.  
 TITLE OF INVENTION: POLYNUCLEOTIDES COEXPRESSED WITH MATRIX-REMODELING GENES  
 FILE REFERENCE: PB-0004 CIP  
 CURRENT APPLICATION NUMBER: US/09/818,143  
 CURRENT FILING DATE: 2001-03-26  
 NUMBER OF SEQ ID NOS: 23  
 SOFTWARE: PERL Program  
 SEQ ID NO 14  
 LENGTH: 2040  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: 2814981CB1  
 US-09-818-143-14

Alignment Scores:  
 Pred. No.: 1,43e-175 Length: 2040  
 Score: 1635.00 Matches: 325  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 97.38% Indels: 0  
 Gaps: 0  
 DB: 9

US-10-617-443B-2 (1-334) x US-09-818-143-14 (1-2040)

10 GYLeuHieGInLeuSerProArgTyrLysPheAsnPhelLeaAspValValGlu 29  
 72 GGTCTCCACCGTAGAGCAGCCCGCGCTCAAGTTCATTGCTGACGCGGGAG 131  
 30 LysIleAlaProAlaValAlaHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49  
 132 AAGATGCGACACAGCCGCGGTCCACATAGAGCTCTTCTTGAGACACCGCGCTGTTGCCCG 191  
 50 AsnValProLeuSerSerSerGlySerGlyPheIleMetSerGluIleGlyLeuIleThr 69  
 192 AACGTGCCCTGCTCCAGCGGCTTCTGCTCATATGTCAGAGCGCGCGCTGATCATCACC 251  
 70 AsnAlaHisIleValIleSerSerAsnSerAlaAlaProGlyArgGlnGlnValGln 89  
 252 AATGCCACCTGCTGTCTCAGCAACAGTCTGCCCGCGGAGGACGACGCTCAAGGTGCAG 311  
 90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLeuAspIleAspIleValSerAspIle 109  
 312 CTACAGATGGGAGCTCTCTATGAGGCCACCATCAAGACATCAAGAAAGTGGACATT 371  
 110 AlaThrIleLeuValIleHisProGlyLysValLeuProValLeuLeuGlnGlnHisSerAla 129  
 372 GCCACCATCAAGATTCATCCCAAGAAAGCTCTCTGTTGTTGCTGGGTCACTGGGCC 431  
 130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149  
 432 GACCTGGGCGCTGGGAGATTGTGTGGCCATGGCAGTCCCTTCCGCTCAAGAACACA 491  
 150 ValThrThrGlyIleValIleSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAsp 169  
 492 GTACAAACGGGACCTGTCACACACTGCCACGCGGAGGAGGAGGAGCGTGGCCCGGAC 551  
 170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189  
 552 TCCGACATGGACATCACTCAAGACGATGCCATCATCACTACAGGAACCTCCGGGAGACA 611  
 190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuGlyValThrAlaGlyIle 209  
 612 CTGGTGAACCTGGATGGCGAGTCAATGGCATCAACGCTCAAGGCTCAAGCGTGGCATC 671  
 210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspIleValGln 229  
 672 TCCCTTGGCATCCCTCAGACCGGACATCAAGGTTCTCTCAAGAGTTCAAGAACAGCAG 731

230 IleYAspTyrPheLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249  
 732 ATCAAGACTGGAAAGACCGCTTCAATCGCATACCGATCCGAGCATCAACCAAGCCCTG 791  
 250 ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleThrVal 269  
 792 GTGATGAGCTGAAGGACGACCAACCGGACTTCCAGAGGCTCAGACGATGAAATTTATGTC 851  
 270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleIle 289  
 852 CAAGAGGTTGCCCGCAATTCACCTTCTCAGAGAGCGGACATCAAGATGGTACATCTAC 911  
 290 ValIysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluValAlaValLeuThr 309  
 912 GTCAAGGCTCAACGGCGCTCTCTTATGATCTGATGAGCTGCAAGAGCGGTGCTGACCC 971  
 310 GluSerProLeuLeuLeuGluValAlaArgGlyAsnAspAspLeuPheSerIleAla 329  
 972 GAGTCTCCTCTCTCACTGAGAGGTGCGGCGGGAACGACGACTCTTCAAGCATCGCA 1031  
 330 ProGluValAlaValMet 334  
 1032 CCTGAGGTGTCATG 1046

RESULT 7

US-10-381-820A-7  
 Sequence 7, Application US/10381820A  
 Publication No. US20040142333A1

GENERAL INFORMATION:  
 APPLICANT: Deshun, Lu  
 APPLICANT: Song, Ho Yeong  
 APPLICANT: Su, Eric Wen  
 APPLICANT: Wang, He  
 TITLE OF INVENTION: Novel Secreted Proteins and Their Uses  
 FILE REFERENCE: X-13974  
 CURRENT APPLICATION NUMBER: US/10/381,820A  
 CURRENT FILING DATE: 2003-03-27  
 NUMBER OF SEQ ID NOS: 8  
 SOFTWARE: PatentIn version 3.2  
 SEQ ID NO 7  
 LENGTH: 2094  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (206) ..(1564)  
 US-10-381-820A-7

Alignment Scores:  
 Pred. No.: 1.49e-175 Length: 2094  
 Score: 1635.00 Matches: 325  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 97.38% Indels: 0  
 Gaps: 0  
 DB: 19

US-10-617-443B-2 (1-334) x US-10-381-820A-7 (1-2094)

10 GYLeuHieGInLeuSerProArgTyrLysPheAsnPhelLeaAspValValGlu 29  
 590 GGTCTCCACCGTAGAGCAGCCCGCGCTCAAGTTCATTGCTGACGCGGGAG 649  
 30 LysIleAlaProAlaValAlaHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49  
 650 AAGATGCGACACAGCGGTGCTCAATAGAGCTTCTCTGAGACACCGCGTGTGGCCGC 709  
 50 AsnValProLeuSerSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleThr 69  
 710 AACGTGCCCTGCTCCAGCGGTTCTGCTCATCATGTCAAGAGCGCGCGCTGATCATCACC 769  
 70 AsnAlaHisIleValIleSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuValGln 89  
 770 AATGCCACCTGCTGTCTCAGCAACAGTGTGCCCGGAGGACGACGCTCAAGGTGCAG 829

Qy 90 LeuGlnaengLYAspSerTYrGluAlaThrIleYAspIleAspLYeSerAspIle 109  
Db 830 CTACAGAAATGGGGAATCTCTATAGAGCCACATCAAGACATCGACAAAGATCGGACATT 889  
Qy 110 AlaThrIleYsIleHisProLYsLYsLeuProValLeuLeuGluYHisSerAla 129  
Db 890 GCCACATCAAGATTCATCCCAAGAAAAGCTCCCTGTTGTTGCTGGGTCACTCGGCC 949  
Qy 130 AspleuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149  
Db 950 GACCTGGCGCTGGGAGATTGTGGTGGCCATCGGCACTCCCTTCGCTCAGAACCA 1009  
Qy 150 ValThrThrGlyIleValSerThrAlaGlnArgGluYArgGluLeuGlyLeuAlaArgAsp 169  
Db 1010 GTGACAAAGGCGCATCTGACACATCCCAAGCGGAGGCGAAGAGCTGGGCTCGGAGC 1069  
Qy 170 SerAspMetAspTYrIleGlnThrAspAlaIleIleAsnTYrGlyAsnSerGlyPro 189  
Db 1070 TCCGACATGAGCTACATCCAGACGATCGCATCACTCAAGGAACTCCGGGGAGACA 1129  
Qy 190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLYsValThrAlaGlyIle 209  
Db 1130 CTGGTGAACCTGATGGCGAGGTCAATGGCATCAACGCTCAAGGTCAAGCTGGCATC 1189  
Qy 210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLYsGln 229  
Db 1190 TCTTTTGCCATCCCTCAGACCGCATCACAGGTTCTCCACAGAGTTCCAAAGACAGCAG 1249  
Qy 230 IleLYsAspTrpLYsLYsArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249  
Db 1250 ATCAAGACATGAGAGAGCGCTTCATCGGCATACGAGATCGGACATCAACCAAGCTTG 1309  
Qy 250 ValAspGluLeuLYsAlaSerAsnProAspPheProGluValSerSerGlyIleTYrVal 269  
Db 1310 GTGGATAGCTGAGAGCGCAGAACCCCGGACTTCCAGAGGTCAAGAGTGAATTATGTCG 1369  
Qy 270 GlnGluValAlaProAsnSerProSerGlnArgIleGlyIleGlnAspGlyAspIleIle 289  
Db 1370 CAAGAGGTGGCGGAATTCCTCTTCAAGAGCGGCGCATCAAGAGTGGATCATCTC 1429  
Qy 290 ValLYsValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThr 309  
Db 1430 GTCAAGGTCAAGCGGCGCTCTCTAGTGAATCGAGTGAAGCTGCGAGGCGGCTGTAAC 1489  
Qy 310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuPheSerIleAla 329  
Db 1490 GAGTCTCTCTCTTACTGAGAGGTGCGGGGGGAAAGAGCAAGCACTCTTCAAGATCGCA 1549  
Qy 330 ProGluValValMet 334  
Db 1550 CCTGAGGTGTCATG 1564

RESULT 8  
US-10-956-157-5181  
; Sequence 5181, Application US/10956157  
; Publication No. US20050118625A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956,157  
; NUMBER OF SEQ ID NOS: 319805  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 5181  
; LENGTH: 2541  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-956-157-5181

Alignment Scores:  
Pred. No.: 1,97e-175 Length: 2541  
Score: 1635.00 Matches: 325  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 97.38% Indels: 0  
DB: 21 Gaps: 0  
US-10-617-443b-2 (1-334) x US-10-956-157-5181 (1-2541)

Qy 10 GLYLeuHisGlnLeuSerSerProArgTYrLYsPheAsnPheIleAlaAspValValGlu 29  
Db 589 GGTCTCCACAGCTGAGACAGCCGCGCTTCAAGATTCACTTCAATTCATGAGTGAG 648  
Qy 30 LYsIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49  
Db 649 AAGATCCACAGCGCGGTGCCATGATGACTCTTCTTGAGAACCCGCTGTTGGCGGC 708  
Qy 50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69  
Db 709 AACGTGCCCTCTTCACAGCGTTCTGCTTCATCATGTCAGAGCGCGCTGATCATCAC 768  
Qy 70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLYsValGln 89  
Db 769 AATGCCACAGGTGTCAGACACAGTGTGCCCGGCGAGGCGAGCATCAAGTGCAG 828  
Qy 90 LeuGlnaengLYAspSerTYrGluAlaThrIleLYsAspIleAspLYeSerAspIle 109  
Db 829 CTACAGAAATGGGGAATCTCTATGAGGCCACATCAAAACATCGCAAGAAATCGGACATT 888  
Qy 110 AlaThrIleYsIleHisProLYsLYsArgPheIleGlySerProPheAlaLeuGlnAsnThr 129  
Db 889 GCCACATCAAGATCAATCCCAAGAAAAGTCTCTGTGTGTGCTGGGTCACTCGGCC 948  
Qy 130 AspleuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149  
Db 949 GACCTGGCGCTGGGAGATTGTGGTGGCATCGGCACTCCCTTCGCTCAGAACCA 1008  
Qy 150 ValThrThrGlyIleValSerThrAlaGlnArgGluLYsArgGluLeuArgAsp 169  
Db 1009 GTGCAACGGGATCGTCACACATCGCCAGCGGAGGCGAGGAGCTCGGCGGAGC 1068  
Qy 170 SerAspMetAspTYrIleGlnThrAspAlaIleIleAsnTYrGlyAsnSerGlyPro 189  
Db 1069 TCCGACATGAGCATCAATCCAGACGAGGATCATCATCACTCAAGGAACTCCGGGAGCA 1128  
Qy 190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLYsValThrAlaGlyIle 209  
Db 1129 CTGGTGAACCTGATGGCGAGGTCAATGGCATCAACGCTCAAGGTCAAGCTGCGATC 1188  
Qy 210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLYsGln 229  
Db 1189 TCTTTTGCCATCCCTCAGACCGCATCAACGTTCTCTCAAGAGTTCCAAAGCAAGCAG 1248  
Qy 230 IleLYsAspTrpLYsLYsArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249  
Db 1249 ATCAAGACATGAGAGAGCGCTTCAATCGGATCGGATGGGAGCATCAAGCAAGCTTG 1308  
Qy 250 ValAspGluLeuLYsAlaSerAsnProAspPheProGluValSerSerGlyIleTYrVal 269  
Db 1309 GTGATGAGCTGAAGGCGCAGACACCGGACTTCCAGAGGTCAAGCTGAGAAATTATGTG 1368  
Qy 270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleIle 289  
Db 1369 CAAGAGGTGGCGGAATTCATCTTCAAGAGGCGGCGATCAAGATGTGATCATATC 1428  
Qy 290 ValLYsValAsnGlyArgProLeuValAspSerSerGlyLeuGlnGluAlaValLeuThr 309  
Db 1429 GTCAAGGTCAAGCGGCGTCTCTTACTGAGACTCGAGTGAAGCTGAGAGGCGGTGAGACC 1488  
Qy 310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuPheSerIleAla 329  
Db 1489 GAGTCTCTCTCTTACTGAGAGGTGCGGGGGGAAAGAGCAAGCACTCTTCAAGATCGCA 1548

QY 330 ProgluValValMet 334  
DB 1549 CCTGAGGTGGTCAATG 1563

## RESULT 9

US-10-485-313A-31  
Sequence 31, Application US/10485313A  
Publication No. US2005005902A1  
GENERAL INFORMATION:  
APPLICANT: NTE, Guyling  
APPLICANT: SALAMONSEN, Lois Adrienne  
APPLICANT: LI, Ying  
APPLICANT: HAMPTON, Anne Lorraine  
APPLICANT: FINDLAY, John Kerr  
TITLE OF INVENTION: Novel Serine Protease  
FILE REFERENCE: 31633-200357  
CURRENT APPLICATION NUMBER: US/10/485,313A  
CURRENT FILING DATE: 2004-01-30  
PRIOR APPLICATION NUMBER: PCT/AU02/01010  
PRIOR FILING DATE: 2002-07-30  
PRIOR APPLICATION NUMBER: PR6707  
PRIOR FILING DATE: 2001-07-30  
NUMBER OF SEQ ID NOS: 55  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 31  
LENGTH: 2543  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-485-313A-31

## Alignment Scores:

pred. No.:	1,97e-175	Length:	2543
Score:	1635.00	Matches:	325
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.38%	Indels:	0
DB:	21	Gaps:	0

US-10-617-443B-2 (1-334) x US-10-485-313A-31 (1-2543)

QY 10 GtyleuHieGlnleuSerSerProArgTYrlyPheAsnPhelIeAaSpValValGlu 29  
DB 577 GGCTTCACACAGTGAAGAGCCGCCCTCAATTCACTTCACTTCTGACGCGTGGAG 636  
QY 30 LyvIleAlaProAlaValValHieGlnleuPheleuArgHieProleuPheGlyArg 49  
DB 637 AAGATGCAACAGCCGTGGTCCATAGAGCTCTCTGAGACACCGCGTGTGGCCG 696  
QY 50 AaValProleuSerSerGlySerGlyPheIleMetSerGluIaGlyleuIleIleThr 69  
DB 697 AACGTCCCTCTGTCAGCGGTTCTGCTTCATCATGTCAAGGCCCGGCTGATCATCAC 756  
QY 70 AsnAlaHieValValSerSerAenSerAlaAlaProGlyArgGlnleuIleValGln 89  
DB 757 AATGCCAGGTGGTGTCCAGCAACAGTGTGCGCCGAGGAGGAGCAGCAGTGCAG 816  
QY 90 LeuGlnAsnGlyAaSpSerTYrGluAlaThrIleLysAspIleAspLysLysSerAapIle 109  
DB 817 CTACAGAAATGGGACCTCTATGAGGCCACCATAAAGACATCAAGAAAGTCGACATT 876  
QY 110 AlaThrIleLysIleHieProLysLysLysLeuProValIleuIleuIleuIleHieSerAla 129  
DB 877 GCCACCATCAAGATCCATCCAGAAAGAGCTCCGTGTGTGTGCTGGGTCACTGGCC 936  
QY 130 AspLeuArgProGlyGluPheValAlaAlaIleGlySerProPheAlaLeuGlnAenThr 149  
DB 937 GACCTCGGCGCTGGGAGTTGTGTGGTGGCATGGCAGTCCCTTCCGCCACAGAACACA 996  
QY 150 ValThrThyGlyIleValSerThrAlaGlnArgGlyAaArgGluLeuGlyLeuArgAap 169  
DB 997 GTGACAAACGGGAGCTGTCAAGACTGCCAGCGGAGGAGGAGGAGGAGGAGGAGGAG 1056

QY 170 SerAaMetAspTYrIleGlnThrAspAlaIleIleAenTYrGlyAaSpSerGlyGlyPro 189  
DB 1057 TCCGACATGACACTATCATCAAGCGGATCCATCATCAATCAAGGAGACTCCGGGGAGCA 1116  
QY 190 LeuValAaenLeuAaPglyGluValIleGlyIleAenThrLeuLysValThrAlaGlyIle 209  
DB 1117 CTGGTGAACCTGGAGGAGGAGGTCAATGGCATCAACAGCTCAAGGTCAAGGCTGGCATC 1176  
QY 210 SerPheAlaIleProSerAaPArgIleThrArgPheLeuThrGluPheGlnAaPlyGln 229  
DB 1177 TCCCTTCCCATCCCTCTCAAGCCGATCAAGGTTCTCTCAAGATTCAGACAGACAG 1236  
QY 230 IleLysAaPTryLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249  
DB 1237 ATCAAAAGACTGGAAAGAGGCTTATCGGATAGGATGCGGACATCAACAGACCTG 1296  
QY 250 ValAaPglyLeuLysAlaSerAaProAaPheProGluValSerSerGlyIleTYrVal 269  
DB 1297 GTGATGAGCTGMAAGCCAGCAACCGGACTTCCAGAGGTCAAGCATGGAATTTATGTG 1356  
QY 270 GlnGluValAlaProAaSerProSerGlnArgGlyIleGlnAaPglyAaPlyIle 289  
DB 1357 CAAGAGTTGGCCGCAATTCATCTTCAAGAGGCGGATCAAGATGTCATCATCTC 1416  
QY 290 ValLysValAaenGlyArgProleuValAaSpSerSerGlyLeuGlnGluAlaValleuThr 309  
DB 1417 GTCAAGGTCAACGGCGCTCTCTAGTGAATCTGAGTGAAGGAGGAGGAGGAGGAGGAG 1476  
QY 310 GluSerProleuLeuLeuGluValArgArgGlyAaenAaPheLeuPheSerIleAla 329  
DB 1477 GAGTCTCTCTCTCTAGAGGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1536  
QY 330 ProgluValValMet 334  
DB 1537 CCTGAGGTGGTCAATG 1551

## RESULT 10

US-09-796-753-31  
Sequence 31, Application US/09796753  
Publication No. US20030027998A1  
GENERAL INFORMATION:  
APPLICANT: McCarthy, Sean A.  
TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF  
FILE REFERENCE: 7853-227-999  
CURRENT APPLICATION NUMBER: US/09/796,753  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 09/183,175  
PRIOR FILING DATE: 1998-10-30  
PRIOR APPLICATION NUMBER: 09/223,094  
PRIOR FILING DATE: 1998-12-30  
PRIOR APPLICATION NUMBER: 09/223,546  
PRIOR FILING DATE: 1998-12-30  
PRIOR APPLICATION NUMBER: 09/224,246  
PRIOR FILING DATE: 1998-12-30  
PRIOR APPLICATION NUMBER: 09/259,388  
PRIOR FILING DATE: 1999-02-26  
PRIOR APPLICATION NUMBER: 60/122,458  
PRIOR FILING DATE: 1999-03-01  
PRIOR APPLICATION NUMBER: 09/312,359  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 09/336,536  
PRIOR FILING DATE: 1999-06-18  
PRIOR APPLICATION NUMBER: 09/342,687  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: 09/345,464  
PRIOR FILING DATE: 1999-06-30  
PRIOR APPLICATION NUMBER: 09/365,164  
PRIOR FILING DATE: 1999-07-30  
PRIOR APPLICATION NUMBER: 09/399,723  
PRIOR FILING DATE: 1999-09-20  
PRIOR APPLICATION NUMBER: 09/409,634  
PRIOR FILING DATE: 1999-09-30  
PRIOR APPLICATION NUMBER: 09/471,179





QY 9 AlaGlyLeuHISGlnLeuSerProArgTyrLysPheAsnPheIleAlaAspValVal 28  
Db 526 TCTGGCTCCACAGCTGACGATCCGCGCTACAGTTCACTTCACGATGCGAG 585  
QY 29 GlnValIleAlaProIleValValHisIleGlnLeuPheLeuArgHisProLeuPheGly 48  
Db 586 GAGAAGATTGCGCACCTGTGTTCACATAGAGCTTTCTTGAGACACCCCTGTTGGC 645  
QY 49 ArgAsnValProLeuSerGlySerGlyPheIleMetSerGlnAlaGlyLeuIleIle 68  
Db 646 CGGAATGTCCTGCTCCAGTGGCTCGGGCTTCATCATGCAAGACCGGTTGATCGTC 705  
QY 69 ThrAsnAlaHisValValSerSerAsnSerAlaIleProGlyArgGlnGlnLeuVal 88  
Db 706 ACCAAGCCGACCTGCTTCACAGCTCCAGCATCTCCGCGCGGACGAGCTGAAGGTG 765  
QY 89 GlnLeuGlnAsnGlyAspSerTyrGlnAlaThrIleLeuAspIleAspLysLysSerAsp 108  
Db 766 CAGCTCAGATGGAGATGCTTATGAGGCCACATCCAGACATCCAGAAAGTGGAGC 825  
QY 109 IleAlaThrIleLysIleHisProLysLysLeuProValLeuLeuLeuGlyHisSer 128  
Db 826 ATGGCCAGCATTTGTAATCCACCCCAAGAAAAGCTCCCTGTGTGCTGGCTGCTCACTA 885  
QY 129 AlaAspLeuArgProGlyGluPheValValAlaIleGlySerProPheIleLeuGlnAn 148  
Db 886 GCAGACCTGGCGCTGGAGTTCGTGGTGGCCATGGCAGCCCTTGGCTTGCAGAAC 945  
QY 149 ThrValThrIleGlyIleValSerThrAlaGlnArgLysGluArgLysLeuArg 168  
Db 946 ACCGTGACAAAGGGCATTTGTCACAGCTCCAGCGGGATGGCAAGAGTGGGTCTCCGG 1005  
QY 169 AspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGly 188  
Db 1006 GACTCAGACATGACTATATCCAGACCGATCATCATTAACGGGAAGTCCAGAGGA 1065  
QY 189 ProLeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGly 208  
Db 1066 CCCCTGGTGAACCTGATGGAGGTCATGGCATCAACCGCTCAAGTGGCAGCTGGC 1125  
QY 209 IleSerPheAlaIleProSerAspArgIleThrArgPheLeuThrGlnAspLys 228  
Db 1126 ATCTCTCTTCCATCCCTCAGATCGCATCAACGCTTCTCTGTGATGCCAAACAAAG 1185  
QY 229 GlnIleLysAspTyrLysLysArgPheIleGlyIleArgMetArgThrIleThrProSer 248  
Db 1186 CATGTGAAGACTGGAAGAGACGCTTCATTTGGCATTCGGATGGCGAACATCAAGCCAAAGT 1245  
QY 249 LeuValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyr 268  
Db 1246 TTGGTGGAGAACTGAAGGCGCCACCAAGACTTTCAGCGGTCCAGAGTGAATATAT 1305  
QY 269 ValGlnGlnValAlaProAsnSerProSerGlnArgLysIleGlnAspGlyAspIle 288  
Db 1306 GTTCAAGAGGTGTTCCCAATTCATCTTCAGAGAGAGAGCATCCAAATGGGACATC 1365  
QY 289 IleValIleValAlaGlnGlyArgProLeuValAspSerSerGlyLeuGlnGlnValValLeu 308  
Db 1366 ATCGTCAAAAGTCAATGGCCCGCTCGCGATTCAGCGAGCTGACGAGGCGAGTCTTG 1425  
QY 309 ThrGlnSerProLeuLeuGlnValArgArgLysAsnAspAspLeuLeuPheSerIle 328  
Db 1426 AACGAGCTTTCATCTCTGTGGAGGTGGCGAGGCAATGATATCTCTCTTCAGCATC 1485  
QY 329 AlaProGluValValMet 334  
Db 1486 ATCCCTGAGGTGTCATG 1503

RESULT 12  
US-10-301-822-78  
; Sequence 78, Application US/10301822  
; Publication No. US20030148410A1  
; GENERAL INFORMATION:

APPLICANT: Millennium Pharmaceuticals, Inc.  
APPLICANT: Berger, Allison  
APPLICANT: Guillemette, Tracy L.  
APPLICANT: Kamatkar, Shubhangi  
APPLICANT: Schlegel, Robert  
APPLICANT: Monahan, John E.  
APPLICANT: Thibodeau, Stephen N.  
APPLICANT: Burgart, Lawrence J.  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
THERAPY OF COLON CANCER  
FILE REFERENCE: MPM01-0292RNM  
CURRENT APPLICATION NUMBER: US/10/301,822  
PRIORITY FILING DATE: 2002-11-21  
PRIORITY FILING DATE: 2001-12-10  
PRIORITY FILING DATE: 2001-12-10  
PRIORITY FILING DATE: 2002-03-05  
PRIORITY FILING DATE: 2002-03-05  
PRIORITY FILING DATE: 2002-05-20  
NUMBER OF SEQ ID NOS: 228  
SOFTWARE: FASTSeq for Windows Version 4.0  
SEQ ID NO 78  
LENGTH: 2551  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (222) ... (1580)  
US-10-301-822-78

## Alignment Scores:

Pred. No.:	4,71e-158	Length:	2551
Score:	1482.50	Matches:	319
Percent Similarity:	97.55%	Conservative:	0
Best Local Similarity:	97.55%	Mismatches:	6
Query Match:	88.30%	Indels:	7
DB:	15	Gaps:	0

US-10-617-443B-2 (1-334) x US-10-301-822-78 (1-2551)

QY 10 GlnLeuHISGlnLeuSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29  
Db 606 GGTCTCCACAGCTGAGCAGCCCGCGCTACAGTTCACTTCATGTCGATGGTGGAG 665  
QY 30 LysIleAlaProIleValValHisIleGlnLeuPheLeuArgHisProLeuPheGlyArg 49  
Db 666 AGATTCGACCAAGCGGTGTCATAGAGCTTTCGAGACACCCGCTGTTGGCGC 725  
QY 50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGlnAlaGlyLeuIleIleThr 69  
Db 726 AACGTCCCTCTGCTCCAGCGGTTCTGGCTTCATCATGTCAAGAGCGGCTGATCAACC 785  
QY 70 AsnAlaHisValValSerSerAsnSerAlaIleProGlyArgGlnGlnLeuValValLeu 89  
Db 786 AATGCCACAGGTGTTCAGCAACAGTGTGCGCCGCGGACAGCAGCATCAAGGTGCG 845  
QY 90 LeuGlnAsnGlyAspSerTyrGlnAlaThrIleLysAspIleAspLysLysSerAspIle 109  
Db 846 CTACAGAAATGGAGCTTCTATAGAGCCACATCAAGACATGACAAAGATCGGACAT 905  
QY 110 AlaThrIleLysIleHisProLysLysLeuProValLeuLeuLeuGlyHisSerAla 129  
Db 906 GCACACATCAAGATTCATCCCAAGAAAGCTCCCTGTGTGCTGCTGATCACTCGGCC 965  
QY 130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheIleLeuGlnAsnThr 149  
Db 966 GACCTGCGGT-GGGAGATTTGTGGTGGCATGGAGT-CCCTTGCGCTTACAGAAACA 1023  
QY 150 ValThrThrGlyIleValSerThrAlaGlnArgGlnGlyArgGlyLeuGlyLeuArgAsp 169  
Db 1024 GTACAAAC-GGATTCGT-AGCACTGCCAGCG-GAGGGCAGGGAGCTGGGCTCGGGAGC 1080



Qy 170 SerAspMetAspTyrIle-GlnThrAspAlaIleIle-AsnTyrGlyAsnSerGlyIle 189  
Db 1081 TCCGACATGAGCTACATACAGAGCGATGCCATCTCAACATCGGGAATCCCGGGGAC 1140  
Qy 189 rGluValAsnLeuAspGlyValIleGlyIleAsnThrLeuValThrAlaGlyIle 209  
Db 1141 CACTGCTAACCTGATGCGAGGCTCATGCGATCAACACCTCAAGGTCAAGGCTGCA 1200  
Qy 1201 TCTCTTTGCGATCCCTCCATCAGCGCATCAACGCTTCTCAAGATTCAGACAGC 1260  
Qy 229 InileYAspTrpIleYAspArgPheIleGlyIleArgMetArgThrIleThrProserL 249  
Db 1261 AGATCAAGACTGTAAGAAGCGCTTCATCGGATACGATCGGAGCATCAACCAAGCC 1320  
Qy 249 euValAspGluLeuValAsnSerAspPheProGluValSerSerGlyIleTyrV 269  
Db 1321 TGGTGTGATGACCTAGAGGCCAGCAACCGGACTTCCAGAGTCCAGCATGGAATTTATG 1380  
Qy 269 alGlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleI 289  
Db 1381 TGCAGAGGTGCGCGGATTCACCTTCTCAGAGAGCGCGCATCAAGATGTGATCATCA 1440  
Qy 289 leValIyValAsnGlyArgProLeuValAspSerSerGluLeuGlnIleValIleut 309  
Db 1441 TCGTCAGGTGCAAGCGGCTCTCTAGTGAATCGAGTGAAGTGCAGAGGCGCTGCTCA 1500  
Qy 309 hrGluSerProLeuLeuLeuGlnValArgArgGlyAsnAspAspLeuLeuPheSerIleA 329  
Db 1501 CCGAGTCT 1560  
Qy 329 laProGluValIleMet 334  
Db 1561 CACCTGAGGTGTGATG 1577

## RESULT 13

US-09-764-898-112  
; Sequence 112, Application US/09764898  
; Patent No. US20020090673A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P1201  
; CURRENT APPLICATION NUMBER: US/09/764, 898  
; PRIORITY FILING DATE: 2001-01-17  
; PRIOR APPLICATION DATA REMOVED - consult PAM or file wrapper  
; NUMBER OF SEQ ID NOS: 311  
; SOFTWARE: Patent Ver. 2.0  
; SEQ ID NO 112  
; LENGTH: 1868  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-898-112

## Alignment Scores:

Pred. No.: 2,42e-146 Length: 1868  
Score: 1378.00 Matches: 275  
Percent Similarity: 98.57% Conservative: 0  
Best Local Similarity: 98.57% Mismatches: 4  
Query Match: 82.07% Indels: 0  
Gaps: 0

US-10-617-443b-2 (1-334) x US-09-764-898-112 (1-1868)

Qy 56 GlYSerGlyPheIleMetSerGluValAGlyLeuIleIleThrAsnAlaIleValIleSer 75  
Db 23 GGTTCCTGGCTTCATCATGTGAGAGCGGCTGATCATCAATGCCACGTGTCTCC 82  
Qy 76 SerAsnSerAlaAlaProGlyArgGlnGlnLeuValGlnLeuGlnAsnGlyAspSer 95  
Db 83 AGCAACAGTCTGTCGCCGCGGAGCAGACGCTCAAGGTGAGTAAGAAATGGGACTCC 142

Qy 96 TyrGluAlaThrIleLeuAspIleAspIleAspIleSerAspIleAlaThrIleLeuIleHis 115  
Db 143 TATAGGCGCACATCAAAAGCATGAGAAAGTGGACATTTGCACCATCAAGATCAT 202  
Qy 116 ProIyValYLeuLeuProValLeuLeuGlyIleHisSerAlaAspLeuArgProGlyIle 135  
Db 203 CCCAAAGAAAAGCTCCCTGTGTTGTTCTGCTGATCATCGGCGACCTGCGGCGGAG 262  
Qy 136 PheValValAlaIleGlySerProPheAlaIleGlnIleAsnThrValIleThrThyIleVal 155  
Db 263 TTTGTGTGGCATCGGAGTCCCTTGCTTACAGAACAGAGTGAAGGAGCATGCTC 322  
Qy 156 SerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyrIle 175  
Db 323 AGCATGCCAGCGGAGGCGGAGGAGGAGTGGGCTCCGGAATCCGACATGATCATC 382  
Qy 176 GlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyIleProLeuValAsnLeuAspGly 195  
Db 383 CAGACGATGCGATCATCACTACAGGAACTCCGGGGAGCACTGTGAACCTGATGCG 442  
Qy 196 GluValIleGlyIleAsnThrLeuValThrAlaGlyIleSerPheAlaIleProSer 215  
Db 443 GAGGTCAATTGGCATCAACGCTCAAGTCAAGCTGATCTCTTGTCCATCCCTCA 502  
Qy 216 AspArgIleThrArgPheLeuThrGluPheGlnAspIleGlnIleYAspTrpIleYAsp 235  
Db 503 GACGCGATCACACGCTTCTCAGAGATTCAGACAGACAGATCAAGACTGAGAGAG 562  
Qy 236 ArgPheIleGlyIleArgMetArgThrIleThrProSerLeuValAspGluLeuVal 255  
Db 563 CGCTTCAATCGGCAATACGATGCGGAGCATCAACCAAGCTGATGAGTGAAGGCC 622  
Qy 256 SerAsnProAspPheProGluValSerSerGlyIleTyrValGlnGluValAlaProAsn 275  
Db 623 AGCAACCGGAGCTTCCAGAGGTGAGAGTGAATTTATGTGCAAGAGGTGCCCAAT 682  
Qy 276 SerProSerGlnArgGlyIleGlnAspGlyAspIleIleValIyValAsnGlyArg 295  
Db 683 TCACCTTCTCAGAGAGCGGATTCAGAGATGATGATCATCTCTCAAGGTCAACGGCGCT 742  
Qy 296 ProLeuValAspSerSerGluLeuGlnIleValIleValIleThrGluSerProLeuLeu 315  
Db 743 CTTCTAGTGAATCGAGTGAAGTGAAGGAGGCGGTGTCACCGAGTCTCTCTCTCTCT 802  
Qy 316 GluValArgArgGlyAsnAspAspLeuPheSerIleAlaProGluValIleMet 334  
Db 803 GAGGTGCGGCGGAGGAG 859

## RESULT 14

US-09-969-384-7  
; Sequence 7, Application US/09969384  
; Publication No. US20020192749A1  
; GENERAL INFORMATION:  
; APPLICANT: Moore, et al.  
; TITLE OF INVENTION: Human Gene Polynucleotides, Polypeptides, and Antibodies  
; FILE REFERENCE: PRO55P1  
; CURRENT APPLICATION NUMBER: US/09/969, 384  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: PCT/US01/10542  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/236,384  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/194,118  
; PRIOR FILING DATE: 2000-04-03  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: Patent Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 1695  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-969-384-7

## Alignment Scores:

Pred. No.: 3,11e-117 Length: 1695  
 Score: 1122.00 Matches: 243  
 Percent Similarity: 80.25% Conservative: 13  
 Best Local Similarity: 76.18% Mismatches: 33  
 Query Match: 66.83% Indels: 30  
 DB: 9 Gaps: 6

US-10-617-443B-2 (1-334) x US-09-969-384-7 (1-1695)

QY 10 G1YLEUHIAG1LEUSeSerProArq1rlyPheanPhe11eAlaAPValValGlu 29  
 Db 229 GGTCTCCACAGCTGAGAGCCGCGCTTCAAGTTCACTTCTGCTGAGG 288  
 QY 30 Lys11eAlaProAlaValValHis11eGluLeuPheLeuArgHisProLeuPheGlyArg 49  
 Db 289 AAGATCGCACACCGCGGTGTCACATAGAGCTTCTTCTGAGACACCGCGTGTGGCCGC 348  
 QY 50 AsnValProLeuSeSerGlySerGlyPhe11eMetSerGluAlaGlyLeu11e11eThr 69  
 Db 349 AACGTGCGCTGTCAGCGGTTCTGCTTCATCATGTCAGAGCGCGCTGATCATCACCC 408  
 QY 70 AsnAlaHisValValSerSerAnSerAlaAlaProGlyValGluGluLeuValGln 89  
 Db 409 AATGCCCGCGGTGTCAGACACAGTGTGCGCGCGGAGGAGAGCTCAAGTGCAG 468  
 QY 90 LeuGlnAnGlyAspSerTyrgLualAthr11eLysAsp11eAspLysSerAsp11e 109  
 Db 469 CTACAGAAATGGGACTCTCTATGAGCCACCATCAAGATCAAGAGTGGACATTT 528  
 QY 110 AlaThr11eLys11eHisProLysLysLeuProValLeuLeuLeuGlyHisSerAla 129  
 Db 529 GCCACATCAAGATTCATCCCAAGAAAAGCTCTCTGTGTGTGCTGGGCTCTGGCC 588  
 QY 130 AspLeuArgProGlyGluPheValValAla11eGlySerProPheAlaLeuGlnAnThr 149  
 Db 589 GACCTGGCGCTGGAGACTTGTGTGTGCTGCGCATTCGCGATCTCTTGGCTTACAGAACACA 648  
 QY 150 ValThrThrGly11eValSerThrAlaGlnArgGluGlyValGluLeuArgAsp 169  
 Db 649 GTGACAAAGGCGCATCTGTGACAGCTGCGGAGGAGGAGGAGGAGCTGGGCTGGGAGC 708  
 QY 170 SerAspMetAspTyrg11eGlnThrAspAla11e11eAsnTyrgLysAnSerGlyPro 189  
 Db 709 TCCGACATGAGACTATCATCAGACGAGATCATCATCAACGGAAGTCCGGGAGGAGCA 768  
 QY 190 LeuValAnLeuAspGlyGluVal11eGly11eAnThrLeuLysValThrAlaGly11e 209  
 Db 769 CTGTGAACTGTGATGGGAGGTCTATGGCATCAACGCTCAAGGTCAAGCTGTGGCATC 828  
 QY 210 SerPheAla11eProSerAspArg11eThrArgPheLeuThrGluPheGlnAspLysGln 229  
 Db 829 TCTTTTGGCCATCCCTCAGACCGCATCAACGCTTCTCAAGATTCACAGACAGACAG 888  
 QY 230 11eLysAspTrpLysLysArgPhe11eGly11eArgMetArgThr11eThrProSerLeu 249  
 Db 889 ATCAAA-----GCCCTCCACTG 906  
 QY 250 ---ValAspGluLeuLysAlaSerAnProAspPheProGluValSerSerGly11eLys 268  
 Db 907 GCAGTTTCATTGAGAGAGGGGCTTCTCAGCTTCCCTCTCCATGACCCCTGACGC 966  
 QY 269 ValGlnGluValAlaProAnSerPro-----SerGlnArg 280  
 Db 967 CAGACATGDAACCCAGTGCAGCCAGGCTGTGCTGCTGAGAGGCTGTGCTCAAGAGA 1026  
 QY 281 G1YGLY11eGlnAspGlyAsp11eLeuValLysValAnGlyArgProLeu---ValAsp 299  
 Db 1027 GCTGCTGTGAGAGATCCGCGCATTTCTTCTGTGT-----CCATTAGGAGAGAC 1077  
 QY 300 SerSerGluLeuGlnGluValValLeuThrGlu-----SerProLeuLeu 315  
 Db 1078 AATCTGAGCCAGACAGAGCTGTCTTCCCAAGAAAGCTGAAGTCTTCTCTTGG 1134

RESULT 15  
 US-10-485-313A-32  
 ; Sequence 32, Application US/10485313A  
 ; Publication No. US20050059002A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NIE, Guiying  
 ; APPLICANT: SALAMONSEN, Lois Adrienne  
 ; APPLICANT: LI, Ying  
 ; APPLICANT: HAMPTON, Anne Lorraine  
 ; APPLICANT: FINDLAY, John Kerr  
 ; TITLE OF INVENTION: Novel Serine Protease  
 ; FILE REFERENCE: 31633-200357  
 ; CURRENT APPLICATION NUMBER: US/10/485,313A  
 ; CURRENT FILING DATE: 2004-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/AU02/01010  
 ; PRIOR FILING DATE: 2002-07-30  
 ; PRIOR APPLICATION NUMBER: PR6707  
 ; PRIOR FILING DATE: 2001-07-30  
 ; NUMBER OF SEQ ID NOS: 55  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 32  
 ; LENGTH: 1953  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-485-313A-32

Alignment Scores:  
 Pred. No.: 1,42e-116 Length: 1953  
 Score: 1117.00 Matches: 222  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 66.53% Indels: 0  
 DB: 21 Gaps: 0

US-10-617-443B-2 (1-334) x US-10-485-313A-32 (1-1953)

QY 10 G1YLEUHIAG1LEUSeSerProArq1rlyPheanPhe11eAlaAPValValGlu 29  
 Db 492 GGTCTCCACAGCTGAGAGCCGCGCTTCAAGTTCACTTCTGCTGAGG 551  
 QY 30 Lys11eAlaProAlaValValHis11eGluLeuPheLeuArgHisProLeuPheGlyArg 49  
 Db 552 AAGATCGCACACCGCGGTGTCACATAGAGCTTCTTCTGAGACACCGCGTGTGGCCGC 611  
 QY 50 AsnValProLeuSeSerGlySerGlyPhe11eMetSerGluAlaGlyLeu11e11eThr 69  
 Db 612 AACGTGCGCTGTCAGCGGTTCTGCTTCATCATGTCAGAGCGCGCTGATCATCACCC 671  
 QY 70 AsnAlaHisValValSerSerAnSerAlaAlaProGlyValGluGluLeuValGln 89  
 Db 672 AATGCCCGCGGTGTCAGACACAGTGTGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 731  
 QY 90 LeuGlnAnGlyAspSerTyrgLualAthr11eLysAsp11eAspLysSerAsp11e 109  
 Db 732 CTACAGAAATGGGACTCTCTATGAGCCACCATCAAGATCAAGAGTGGAGACATT 791  
 QY 110 AlaThr11eLys11eHisProLysLysLeuProValLeuLeuLeuGlyHisSerAla 129  
 Db 792 GCCACATCAAGATTCATCCCAAGAAAAGCTCTCTGTGTGTGCTGGGCTGCTGCGCC 851  
 QY 130 AspLeuArgProGlyGluPheValValAla11eGlySerProPheAlaLeuGlnAnThr 149  
 Db 852 GACCTGGCGCTGGAGACTTGTGTGTGCTGCGCATTCGCGATCCCTTCCCTACAGAACACA 911  
 QY 150 ValThrThrGly11eValSerThrAlaGlnArgGluGlyValGluLeuArgAsp 169  
 Db 912 GTGACAAAGGCGCATCTGTGACAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 971  
 QY 170 SerAspMetAspTyrg11eGlnThrAspAla11e11eAsnTyrgLysAnSerGlyPro 189  
 Db 972 TCCGACATGAGACTATCATCAGACGAGTGCATCATCAACGGAAGTCCGGGAGGAGCA 1031  
 QY 190 LeuValAnLeuAspGlyGluVal11eGly11eAnThrLeuLysValThrAlaGly11e 209

Db	1032	CTGGTGAACCTGATGCGAGGTGATGGCATCAACAGCTCAAGTCAAGGCTGCGCATC	1091
Qy	210	SerphealalIeProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln	229
Db	1092	TCCCTTGGCATCCCTCGAGCCGATCACAGGTTCTTCACAGAGTTCCAAGACAAAGCAG	1151
Qy	230	IleLys	231
Db	1152	ATCAA	1157

Search completed: June 30, 2005, 01:51:01  
 Job time : 1067.82 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 29, 2005, 21:07:15 ; Search time 5771.48 Seconds  
(without alignments)  
2202.808 Million cell updates/sec

Title: US-10-617-443B-2  
Perfect score: 1679  
Sequence: 1 MHLLPASPAGHQLSPRYK.....LEVRGNDLLFSIAPEVVM 334

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 6.0 , Ygapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 6847908

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US10617443/runat\_29062005\_135932\_26076/arp\_query.fasta\_1.718  
-DB=EST -QFMT=fastcap -SUPFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10617443.CCGN 1.1\_4385.@runat\_29062005\_135932\_26076 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_g881.\*  
9: gb\_g882.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1528	91.0	2730	3	AK035194 Mus muscu
2	1268	75.5	941	5	BO675502 AGENCOURT
3	1256	74.8	810	6	CD635470
4	1216.5	72.5	1854	3	CR731516
5	1191	70.9	890	5	BO880316
6	1176	70.0	724	6	CD635467
7	1157	68.9	757	7	CN787405
8	1117	66.5	1990	3	CR604057
9	1072.5	63.9	796	7	CO404217

10	1061	63.2	798	7	CV106197
11	1034.5	61.6	973	9	AY406047
12	1034.5	61.6	1655	3	CR590731
13	1034.5	61.6	1911	3	BC011352
14	1034.5	61.6	1931	3	CR603877
15	1034.5	61.6	1951	3	CR624458
16	1034.5	61.6	1992	3	CR623030
17	1034.5	61.6	2002	3	CR605013
18	1034.5	61.6	2005	3	CR593682
19	1034.5	61.6	2035	3	CR597732
20	1031	61.4	792	7	CV106456
21	1030.5	61.4	1753	3	AK090320
22	1026.5	61.1	973	9	AY406049
23	1009	60.1	636	7	CN787739
24	990	59.0	852	7	CK599706
25	964	57.4	1577	3	AK046614
26	961	57.2	881	6	CD518132
27	955.5	56.9	1137	4	BM461040
28	953	56.8	684	5	BQ572908
29	948	56.5	745	7	CO386615
30	945	56.3	1328	3	AK008764
31	943	56.2	616	5	BQ328404
32	936	55.7	589	7	CV348463
33	934	55.6	609	6	CD635468
34	933	55.6	557	9	AY403125
35	904	53.8	650	1	AU296610
36	901.5	53.7	1360	7	CR611805
37	898	53.5	730	7	CK479697
38	892	53.1	557	9	AY403127
39	890	53.0	1603	3	CR590249
40	887	52.8	1494	3	CR601506
41	885.5	52.7	793	7	CN037447
42	882	52.5	1485	3	CR606695
43	882	52.5	1563	3	CR611713
44	882	52.5	1791	3	CR593199
45	880.5	52.4	759	5	BP171264

## ALIGNMENTS

RESULT 1  
AK035194  
LOCUS  
DEFINITION  
AK035194 2730 bp. mRNA linear HTC 03-APR-2004  
Mus musculus 12 days embryo embryonic body between diaphragm region  
and neck cDNA, RIKEN full-length enriched library, clone:9430098F20  
product:hypothetical serine proteases, trypsin family containing  
protein, full insert sequence.

ACCESSION  
AK035194  
VERSION  
AK035194.1 GI:26084456  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus

REFERENCE  
1  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)

REFERENCE  
3  
Garninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE  
4  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

REFERENCE  
5  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

REFERENCE  
6  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

REFERENCE  
7  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

JOURNAL	MEDLINE PUBMED REFERENCE AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
TITLE		5
TITLE		The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II team.
JOURNAL	REFERENCE AUTHORS	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 2730)
AUTHORS		Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Haehizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Horii,F., Imotani,K., Ishii,Y., Itchi,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,D., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Satoh,H., Sakai,C., Sakai,K., Sakazune,N., Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akaira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE		Direct Submission
JOURNAL		Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-resgsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/. Location/Qualifiers
FEATURES	source	1..2730 /organism="Mus musculus" /mol_type="mRNA" /bstraln="CS7BL/6J" /db_xref="FANTOM_DB:9430098P20" /db_xref="taxon:10090" /cclone="9430098P20" /tissue_type="embryonic body between diaphragm region and neck" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="12 days embryo" /note="hypothetical Serine proteases, trypsin family containing protein (Interpro IPR001254, evidence: interpro)
ORIGIN		putative" 2715..2720 /note="putative" 2730 /note="putative"
Alignment Scores:		8.38e-174                  Length:                  2730
Aligned No.:		

Score:	1528.00	Matches:	301
Percent Similarity:	96.32%	Conservative:	12
Best Local Similarity:	92.33%	Mismatch:	13
Query Match:	91.01%	Indels:	0
DB:	3	Gaps:	0
US-10-617-443B-2 (1-334) x AK035194 (1-2730)			
QY	9	AlaGlyLeuHISGlnLeuSerProArgTyrIysPheAsnPheIAlaAspValVal	28
DB	837	TCTGTCCTCCACCAAGCTGACCAAGTCGCGGTACAAATTCACTCATCGCGATGTGCTG	896
QY	29	GluIysIleAlaProAlaValValHisIleGlnLeuPheLeuArgHisProLeuPheIly	48
DB	897	GAGAAAGATTGGCCGACGCTGTGCTCACATTAAGCTCTTTCGAGACACCCCTGTTGGC	956
QY	49	ArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGlnAlaGlyLeuIleIle	68
DB	957	CGGATATGGCCGCTCTCCAGTGGCTCGGGCTTTCATCATGTACGAAGCCGCTTATGCTC	1016
QY	69	ThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyIArgGlnGlnLeuVal	88
DB	1017	ACCAAGCCCAAGTGTCTCCAGCTCCAGCACTGCTCCGCGCGAGCACTGAAGCTG	1076
QY	89	GlnLeuGlnAsnGlyAspSerTyrGlnAlaThrIleIysAspIleAspIysSerIysP	108
DB	1077	CAGCTGCAGATGGGAGATGGCTTAAGGCCACCACTCCAGGACATGCAAAAGATCGAC	1136
QY	109	IleAlaThrIleIysIleHisProIysIleIysLeuProValLeuLeuGlyHisSer	128
DB	1137	ATTGCGACAGATTGTATATCCACCCCAAGAAAAGCTCCCTGTGTGTGCTGTGGTACCTCA	1196
QY	129	AlaAspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsn	148
DB	1197	GCAAGCTCGCGCGCTGGGAGTGTGGTGCCATCGGCAAGCCCTTGGCCCTGACAGAAC	1256
QY	149	ThrValThrThrGlyIleValSerThrAlaGlnArgGluIysArgGluLeuGlyLeuArg	168
DB	1257	ACCGTAGCAAGGAGATTTGTACGACTGCCAGCGGAGTAGCAGAGAGACTGGTCTCCGG	1316
QY	169	AspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGly	188
DB	1317	GACTCAGACATGACTATATCCAGACCGATCCATCATATTCGGGAACTCAGAGAGA	1376
QY	189	ProteinValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuIysValThrAlaGly	208
DB	1377	CCCTGTGGAACTCGATGGGAGGTATCGGCATCAACAGCTCAAGGTTGACGCTGGC	1436
QY	209	IleSerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLys	228
DB	1437	ATCTCTTTTGCATCCCTCTCAGATCGCATCAAGGCTTCTCTGTGATTCGAAACAG	1496
QY	229	GlnIleIysAspTyrIysIysArgPheIleGlyIleArgMetArgThrIleThrProSer	248
DB	1497	CATGTGAAGACTCGAAGAACCGCTTCATTTGATGATCCGGATGGGACATCAGCCCAAGT	1556
QY	249	LeuValAspGluLeuIysAlaSerAsnProAspPheProGluValIleSerSerGlyIleTyr	268
DB	1557	TTGTGTGAGGAACTGAAGGCGCGCAACCACTTTCAGCGGCTCACAGAGGAATATAT	1616
QY	269	ValGlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnAspGlyAspIle	288
DB	1617	GTTCAAGAGGTGTCTCCCATTCCTCTTCAGAGAGAGAGCATCAAGATGGGACATC	1676
QY	289	IleValIysValAsnGlyIArgProLeuValAspSerSerGlnLeuGlnGlnValIleu	308
DB	1677	ATCGTCAAGCAATGCGCGCCCTCGCGGATTCACGAGAGTGCAGAGGCGAGTCTTG	1736
QY	309	ThrGlnSerProLeuLeuGlnValIArgArgGlyAsnAspLeuLeuPheSerIle	328
DB	1737	AACAGAGCTTCACCTCGCTGGAGGTGGCGAGAGCATATGATCTCTTCACGATC	1796
QY	329	AlaProGluValAlaMet	334

Db 1797 ATCCCTGAGTGTGTCATG 1814

RESULT 2  
LOCUS B0675502 941 bp mRNA linear EST 15-JUL-2002  
DEFINITION AGENCOURT 8191926 NIH\_MGC\_102 Homo sapiens CDNA clone IMAGE:6256985  
ACCESSION B0675502  
VERSION B0675502.1 GI:21786536  
KEYWORDS EST.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens (human)  
REFERENCE 1 (bases 1 to 941)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LUCM2410 row: 1 column: 18  
High quality sequence stop: 605.  
Location/Qualifiers  
1..941  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6256985"  
/tissue\_type="epidermoid carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC\_102"  
/note="Organ: salivary gland; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; CDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGACGAG(G). Library constructed  
by Ling Hong in the laboratory of Gerald M. Rubin  
(University of California, Berkeley) using ZAP-CDNA  
synthesis kit (Stratagene) and Superscript II RT (Life  
Technologies). Note: this is a NIH\_MGC library."

ORIGIN

Alignment Scores:

Pred. NO.:	6,61e-143	Length:	941
Score:	1268.00	Matches:	255
Percent Similarity:	95.60%	Conservative:	6
Best Local Similarity:	93.41%	Mismatches:	10
Query Match:	75.52%	Indels:	2
DB:	5	Gaps:	0

US-10-617-443B-2 (1-334) x B0675502 (1-941)

Qy 16 SerProArGTYrLYsPheAsnPhIleAlaAPVaValGluLYsIleAlaProAlaVal 35  
Db 2 AGCCCGCGCTCAAGATTCACTTCCTTCTCACTGCTGAGAGAGATGACACCGCCGTG 61

Qy 36 ValHisIleGluLeuPheLeuArgHisPheLeuPheGlyArgAsnValProLeuSerSer 55  
Db 62 GTCCACATAGAGCTCTTCTCTGAGACACCGCGCTGTGGCCCAACGTGCCCTGTCCAGC 121

Qy 56 GlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThrAsnAlaHisValValSer 75  
Db 122 GGTTCGTGCTTCATCATGTCAAGAGCCGCGCTGATCATCAACCAATGCCACGTGTGCC 181

Qy 76 SerAsnSerAlaAlaProGlyArgGlnGlnLeuLeuValGlnLeuGlnAsnGlyAspSer 95

Db 182 AGCAACAGTGTGCCCCGGGAGGACAGCAGCTCAAGGTGACGTACAGAAATGGGACTCC 241

Qy TyrcGluAlaThrIleLeuAspIleLeuAspLeuSerAspIleAlaThrIleValIleHis 115  
Db 242 TATGAGGCCACCATCAAGACATGCAGAGAGTGGACATGTCACCATCAACATCATCAT 301

Qy 116 ProLYsLYsLYsLeuProValAlleuLeuGluGlyHisSerAlaAspLeuArgProGlyGlu 135  
Db 302 CCCAAGAAAAGACTCCCTGTGTTCTTCTGGGTCACTGGCGGACCTCGCGGCTGGGAG 361

Qy 136 PheValValAlaIleGlySerProPheAlaLeuGlnAsnThrValIleThrGlyIleVal 155  
Db 362 TTGTGTGGCGCATCGGACAGCTCTTGCCCTACAGAACACAGTACACAGCGCATCGTC 421

Qy 156 SerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyrIle 175  
Db 422 AGCACTGCCAGCGGAGGAGGAGGAGCTGGCTCGGAGCTCCGACATGAGACTTACATC 481

Qy 176 GlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeuAspGly 195  
Db 482 CAGCGAGTGCATCATCACTACAGGAACTCCGGGGACCATCTGTGAACCTGATGCG 541

Qy 196 GluValIleGlyIleAsnThrLeuLYsValThrAlaGlyIleSerPheAlaIleProSer 215  
Db 542 GAGGTCAATGGCATCAACACGCTCAAGTCAAGCTGCGATCTCTTGGCATCCCTCA 601

Qy 216 AspArgIleThrArgPheLeuThrGluPheGlnAspGlyGlnIleLYsAspTyrGly 235  
Db 602 GACCGCATCACACGCTTCCTCAACAGATTCCAGACACACAGATCAAGCTGGAANA 661

Qy 235 SarGpHeIleGlyIleArgMetArgThrIleThrProSerLeuValAspGluLeuLYsAl 255  
Db 662 GCCCTTATGCGCATACGATGCGGACGATCACCAAGCTTGTGATGAATCTGAAGG 721

Qy 255 AserAsnProAspPheProGluValAserSerGlyIleTyrValGlnGluValAlaProAs 275  
Db 722 CANCAACCCGAGCTTCCCAAGGTTATCATGTGACATTATGCGAACAGATTGCCCGGAA 781

Qy 275 nSerPro-SerGlnArgGlyGlyIleGlnAspGly 286  
Db 782 TTCACCTTTTCGAAAGGGGCGCATCCAGATGGG 816

RESULT 3  
LOCUS CD635470/c 810 bp mRNA linear EST 12-JAN-2004  
DEFINITION S610066301 FLP Homo sapiens CDNA, mRNA sequence.  
ACCESSION CD635470  
VERSION CD635470.1 GI:40283737  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 810)  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE Fu, G. K., Wang, J. T., Yang, J., Au-Yang, J. and Stuve, L. L.  
Circular rapid amplification of cDNA ends for high-throughput  
extension cloning of partial genes  
Genomics 84 (1), 205-210 (2004)  
COMMENT Contact: Fu GK  
1160 Porter Dr., Palo Alto, CA 94304, USA  
Tel: 6508454102  
Email: gfu@incyte.com

FEATURES  
source  
1..810  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone\_lib="FLP"  
/note="Vector: pDrive Cloning Vector"

ORIGIN

Alignment Scores:

Pred. No.:	1,51e-141	Length:	810
Score:	1256.00	Matches:	225
Percent Similarity:	99.22%	Conservative:	0
Best Local Similarity:	99.22%	Mismatches:	0
Query Match:	74.81%	Indels:	2
DB:	6	Gaps:	0

Accession	Gene	Protein	Length	Score	E-value	Species	Accession	Gene	Protein	Length	Score	E-value	Species
QY	79	Ala1a1Progl1yArpGlnGlnLeuLysVal1GlnLeuGlnAsnGlyAsp-SerTyrGluAl	98				QY	79	Ala1a1Progl1yArpGlnGlnLeuLysVal1GlnLeuGlnAsnGlyAsp-SerTyrGluAl	98			
Db	810	GCTGCCCCGGGCGAGGACAGCTCAAGAGGTCAGCTCAACAAGGGGAGACCTCTATGAGGC	751				Db	810	GCTGCCCCGGGCGAGGACAGCTCAAGAGGTCAGCTCAACAAGGGGAGACCTCTATGAGGC	751			
QY	98	athr-1lelysaep11eap1ylysaSerAsp11ealathr1lelysi1eh1sProlySL	1168				QY	98	athr-1lelysaep11eap1ylysaSerAsp11ealathr1lelysi1eh1sProlySL	1168			
Db	750	CACCCATCAAAAGCATTCGACAGAAAGTGGGACATTGGCCACATCAAGATTCATCCCAAGA	699				Db	750	CACCCATCAAAAGCATTCGACAGAAAGTGGGACATTGGCCACATCAAGATTCATCCCAAGA	699			
QY	118	yalyb1eupProval1leuLeuLeuGlyHisSer11aap1eupArpProgl1yGluPheVal1V	1358				QY	118	yalyb1eupProval1leuLeuLeuGlyHisSer11aap1eupArpProgl1yGluPheVal1V	1358			
Db	690	AAAAGCTCCCTGTGTGTGTCTGGGACATCGGCGGACCTGTGGGCTGGGGAGTTTGGG	633				Db	690	AAAAGCTCCCTGTGTGTGTCTGGGACATCGGCGGACCTGTGGGCTGGGGAGTTTGGG	633			
QY	138	ala1a1lel1e1eas1erProPheala1eug1AsnThr1ValThrThrGly11ealSerThra	158				QY	138	ala1a1lel1e1eas1erProPheala1eug1AsnThr1ValThrThrGly11ealSerThra	158			
Db	630	TGGCCATCGGCACTCCCTTCGCCCTTACAGAACACAGTGAACAGGGGCACTGTACAGCTG	571				Db	630	TGGCCATCGGCACTCCCTTCGCCCTTACAGAACACAGTGAACAGGGGCACTGTACAGCTG	571			
QY	158	laGlna1rg1uG1yArG1uLeuG1yLeuA1yAspSerAspMetAspTyr11eGlnThra	178				QY	158	laGlna1rg1uG1yArG1uLeuG1yLeuA1yAspSerAspMetAspTyr11eGlnThra	178			
Db	570	CCGACGGGGAGGCGAGGAGCTGGGCGCTCCGGGACCTCGACATCGACATGACATCAACAGCG	511				Db	570	CCGACGGGGAGGCGAGGAGCTGGGCGCTCCGGGACCTCGACATCGACATGACATCAACAGCG	511			
QY	178	sp1a1a1e1e1eas1erTyr1yAsnSerG1y1ProleuVal1AsnLeuAsp1yGluVal1	198				QY	178	sp1a1a1e1e1eas1erTyr1yAsnSerG1y1ProleuVal1AsnLeuAsp1yGluVal1	198			
Db	510	ATGCCATCATCAACTACGGGAACCTCGGGGGACCACTGGTGAACCTGGATGCGAGGTCA	451				Db	510	ATGCCATCATCAACTACGGGAACCTCGGGGGACCACTGGTGAACCTGGATGCGAGGTCA	451			
QY	198	leG1y1le1eas1erThreuleuLysVal1ThrAlaG1y1leSerPheAla1leProSerAspArg1	218				QY	198	leG1y1le1eas1erThreuleuLysVal1ThrAlaG1y1leSerPheAla1leProSerAspArg1	218			
Db	450	TTGGCATCAACACAGCTCAAGTCAAGGCTGCGATCTCTTGGCCATCCCTCAGACCGCA	391				Db	450	TTGGCATCAACACAGCTCAAGTCAAGGCTGCGATCTCTTGGCCATCCCTCAGACCGCA	391			
QY	218	1e1h1r1rghPheLeuThrGluPheGln1Asp1yGln11e1yAsp1r1y1y1yArp1e1	238				QY	218	1e1h1r1rghPheLeuThrGluPheGln1Asp1yGln11e1yAsp1r1y1y1yArp1e1	238			
Db	390	TCAACGGGTTCCTCACAGATTCACAGAACACAGATCAAAAGACTGGAAGAGGGCTTCA	331				Db	390	TCAACGGGTTCCTCACAGATTCACAGAACACAGATCAAAAGACTGGAAGAGGGCTTCA	331			
QY	238	1eG1y1le1eArpMetArG1yThr11eThrProSerLeuVal1Asp1yLeuLysVal1AsSerAsp	258				QY	238	1eG1y1le1eArpMetArG1yThr11eThrProSerLeuVal1Asp1yLeuLysVal1AsSerAsp	258			
Db	330	TGGGATTAAGGATGCGGAGATCAACACAGCTGGTGGATGAGCTGAAGGCCAGCAACC	271				Db	330	TGGGATTAAGGATGCGGAGATCAACACAGCTGGTGGATGAGCTGAAGGCCAGCAACC	271			
QY	258	roAspPheProGluVal1SerSerGly11eTyrVal1GlnGluVal1AlaProAsnSerProS	278				QY	258	roAspPheProGluVal1SerSerGly11eTyrVal1GlnGluVal1AlaProAsnSerProS	278			
Db	270	CGGACTTCCGACGGGTCAAGACAGTGAATTATGTCAGAGGTTGGCCGCAATTCACCTT	211				Db	270	CGGACTTCCGACGGGTCAAGACAGTGAATTATGTCAGAGGTTGGCCGCAATTCACCTT	211			
QY	278	erGlna1rg1yG1y11eG1na1spG1yAsp11e1eal1y1eVal1yAsnG1yArpProleuY	298				QY	278	erGlna1rg1yG1y11eG1na1spG1yAsp11e1eal1y1eVal1yAsnG1yArpProleuY	298			
Db	210	CTCAGAGAGGCGGATCATCAAGATGGTACATATGTCATAGGTCAAGGGGGCTCTTAG	151				Db	210	CTCAGAGAGGCGGATCATCAAGATGGTACATATGTCATAGGTCAAGGGGGCTCTTAG	151			
QY	298	a1AspSerSerG1uLeuGlnGluVal1Val1LeuThrG1uSerProleuLeuLeuGluVal1A	318				QY	298	a1AspSerSerG1uLeuGlnGluVal1Val1LeuThrG1uSerProleuLeuLeuGluVal1A	318			
Db	150	TGGACTCGAGTAGCTGCGAGAGGCGGAGCTGACCGAGATCTCTCTACTGAGAGTGC	91				Db	150	TGGACTCGAGTAGCTGCGAGAGGCGGAGCTGACCGAGATCTCTCTACTGAGAGTGC	91			
QY	318	rgArG1yAsnAspAspLeuLeuPheSer11eAlaProGluVal1 333					QY	318	rgArG1yAsnAspAspLeuLeuPheSer11eAlaProGluVal1 333				
Db	90	GGGGGGGGAACAGACGACCTCTTCAGATCGCACCTGAGAGTGATC 44					Db	90	GGGGGGGGAACAGACGACCTCTTCAGATCGCACCTGAGAGTGATC 44				

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Tetradontoidea, Tetradontidae, Tetraodon.  
1 (bases 1 to 1854)  
Genoscope.  
Direct Submission  
Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

COMMENT The sequences are based on single pass reads.

<http://www.genoscope.cns.fr/tetraodon>.

```
source
1. .1854
/organism="Tetradodon nigrorividiu"
```

```
/db xref="taxon:99883"
```

ORIGIN

/s20000\_cjpc- zara

**Alignment Scores:**

Score:	1216.50	Matches:	240
--------	---------	----------	-----

Best Local Similarity:	72.51%	Mismatches:	46
Query Match:	73.45%	Indels:	4

Year	Number of cases	Percentage of cases
1990	10	10.0
1991	15	15.0
1992	20	20.0
1993	25	25.0
1994	30	30.0
1995	35	35.0
1996	40	40.0
1997	45	45.0
1998	50	50.0
1999	55	55.0
2000	60	60.0
2001	65	65.0
2002	70	70.0
2003	75	75.0
2004	80	80.0
2005	85	85.0
2006	90	90.0
2007	95	95.0
2008	100	100.0
2009	105	105.0
2010	110	110.0
2011	115	115.0
2012	120	120.0
2013	125	125.0
2014	130	130.0
2015	135	135.0
2016	140	140.0
2017	145	145.0
2018	150	150.0
2019	155	155.0
2020	160	160.0
2021	165	165.0
2022	170	170.0
2023	175	175.0
2024	180	180.0
2025	185	185.0
2026	190	190.0
2027	195	195.0
2028	200	200.0
2029	205	205.0
2030	210	210.0
2031	215	215.0
2032	220	220.0
2033	225	225.0
2034	230	230.0
2035	235	235.0
2036	240	240.0
2037	245	245.0
2038	250	250.0
2039	255	255.0
2040	260	260.0
2041	265	265.0
2042	270	270.0
2043	275	275.0
2044	280	280.0
2045	285	285.0
2046	290	290.0
2047	295	295.0
2048	300	300.0
2049	305	305.0
2050	310	310.0
2051	315	315.0
2052	320	320.0
2053	325	325.0
2054	330	330.0
2055	335	335.0
2056	340	340.0
2057	345	345.0
2058	350	350.0
2059	355	355.0
2060	360	360.0
2061	365	365.0
2062	370	370.0
2063	375	375.0
2064	380	380.0
2065	385	385.0
2066	390	390.0
2067	395	395.0
2068	400	400.0
2069	405	405.0
2070	410	410.0
2071	415	415.0
2072	420	420.0
2073	425	425.0
2074	430	430.0
2075	435	435.0
2076	440	440.0
2077	445	445.0
2078	450	450.0
2079	455	455.0
2080	460	460.0
2081	465	465.0
2082	470	470.0
2083	475	475.0
2084	480	480.0
2085	485	485.0
2086	490	490.0
2087	495	495.0
2088	500	500.0
2089	505	505.0
2090	510	510.0
2091	515	515.0
2092	520	520.0
2093	525	525.0
2094	530	530.0
2095	535	535.0
2096	540	540.0
2097	545	545.0
2098	550	550.0
2099	555	555.0
2100		

6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529

.....

[illegible]

631 590

47 pHeGIVArqAsnValProLeuSerSerGlvSerGlvPheIleMetSerGluAlaGlvLeu 66

Db 691 TTTGGTGGACTGTCCCCCTGTCAAGTGGATCCGGCTTCTCTGTGTCCGAGAACGGCCTG 750

67 IleIleThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeu 86

Db 751 ATCGTCACCAAGCCACGTCCTCCAGCACCTCCCCCGGACGGCTCAGCAGCACCTC 810

87 LysValGlnLeuGlnAsnGlyAspSerTyrGlnAlaThrIleLysAspIleAspPylLys 106

DD 011 AAGG1CCHG1GCAAGAACGGGGAACHCG1ACGGAAAGCCHGCG1CAGAGACAA1GACAAAGAAAG 8/V

[illegible][illegible]

On 147 c] nAenTherV[a] tTherTherG] vI] aV[a] SoerTherA] aG] nArerG] wG] wArerG] wI] aV[G] v 166

Db 991 CAGAACACCGTCACCAACCGCATCTCAGCAACCGCTCAGCGGACGGCAGAGAGCTGGGC 1050

167 LeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSer 186

Db 1051 CTGAGGACTCGGAACATGCTACGTCAGACGGAGCCATCATCACTACGGCACTCT 1110

187 GLYGLYPROLEUVALASNLEUABPGLYGLUVALILEGLYILEASNTHRLEULYSVALTHR 206

Db 1111 GGAGGGCCCCCTGTCATTTGGACGGAGAGGTGATCGGGATCAACACCCTGAAGTGGCG 1170

20 / MIAgYIieselPNeAlaIlePNeSeIAbPaRgIleInIRAgPNeLeuInr----- 223

[illegible]

1. *Chlorophyll a* (Chl *a*)

2. *Chlorophyll b* (Chl *b*)

3. *Carotenoids* (Car)

4. *Phaeophytin a* (Phe *a*)

5. *Phaeophytin b* (Phe *b*)

6. *Phaeoerythrin* (Phe *e*)

7. *Phaeoxanthophyll* (Phe *x*)

8. *Phaeo-zeaxanthin* (Phe *z*)

9. *Phaeo-zeaxanthin* (Phe *z*)

10. *Phaeo-zeaxanthin* (Phe *z*)

11. *Phaeo-zeaxanthin* (Phe *z*)

12. *Phaeo-zeaxanthin* (Phe *z*)

13. *Phaeo-zeaxanthin* (Phe *z*)

14. *Phaeo-zeaxanthin* (Phe *z*)

15. *Phaeo-zeaxanthin* (Phe *z*)

16. *Phaeo-zeaxanthin* (Phe *z*)

17. *Phaeo-zeaxanthin* (Phe *z*)

18. *Phaeo-zeaxanthin* (Phe *z*)

19. *Phaeo-zeaxanthin* (Phe *z*)

20. *Phaeo-zeaxanthin* (Phe *z*)

21. *Phaeo-zeaxanthin* (Phe *z*)

22. *Phaeo-zeaxanthin* (Phe *z*)

23. *Phaeo-zeaxanthin* (Phe *z*)

24. *Phaeo-zeaxanthin* (Phe *z*)

25. *Phaeo-zeaxanthin* (Phe *z*)

26. *Phaeo-zeaxanthin* (Phe *z*)

27. *Phaeo-zeaxanthin* (Phe *z*)

28. *Phaeo-zeaxanthin* (Phe *z*)

29. *Phaeo-zeaxanthin* (Phe *z*)

30. *Phaeo-zeaxanthin* (Phe *z*)

31. *Phaeo-zeaxanthin* (Phe *z*)

32. *Phaeo-zeaxanthin* (Phe *z*)

33. *Phaeo-zeaxanthin* (Phe *z*)

34. *Phaeo-zeaxanthin* (Phe *z*)

35. *Phaeo-zeaxanthin* (Phe *z*)

36. *Phaeo-zeaxanthin* (Phe *z*)

37. *Phaeo-zeaxanthin* (Phe *z*)

38. *Phaeo-zeaxanthin* (Phe *z*)

39. *Phaeo-zeaxanthin* (Phe *z*)

40. *Phaeo-zeaxanthin* (Phe *z*)

41. *Phaeo-zeaxanthin* (Phe *z*)

42. *Phaeo-zeaxanthin* (Phe *z*)

43. *Phaeo-zeaxanthin* (Phe *z*)

44. *Phaeo-zeaxanthin* (Phe *z*)

45. *Phaeo-zeaxanthin* (Phe *z*)

46. *Phaeo-zeaxanthin* (Phe *z*)

47. *Phaeo-zeaxanthin* (Phe *z*)

48. *Phaeo-zeaxanthin* (Phe *z*)

49. *Phaeo-zeaxanthin* (Phe *z*)

50. *Phaeo-zeaxanthin* (Phe *z*)

51. *Phaeo-zeaxanthin* (Phe *z*)

52. *Phaeo-zeaxanthin* (Phe *z*)

53. *Phaeo-zeaxanthin* (Phe *z*)

54. *Phaeo-zeaxanthin* (Phe *z*)

55. *Phaeo-zeaxanthin* (Phe *z*)

56. *Phaeo-zeaxanthin* (Phe *z*)

57. *Phaeo-zeaxanthin* (Phe *z*)

58. *Phaeo-zeaxanthin* (Phe *z*)

59. *Phaeo-zeaxanthin* (Phe *z*)

60. *Phaeo-zeaxanthin* (Phe *z*)

61. *Phaeo-zeaxanthin* (Phe *z*)

62. *Phaeo-zeaxanthin* (Phe *z*)

63. *Phaeo-zeaxanthin* (Phe *z*)

64. *Phaeo-zeaxanthin* (Phe *z*)

65. *Phaeo-zeaxanthin* (Phe *z*)

66. *Phaeo-zeaxanthin* (Phe *z*)

67. *Phaeo-zeaxanthin* (Phe *z*)

68. *Phaeo-zeaxanthin* (Phe *z*)

69. *Phaeo-zeaxanthin* (Phe *z*)

70. *Phaeo-zeaxanthin* (Phe *z*)

71. *Phaeo-zeaxanthin* (Phe *z*)

72. *Phaeo-zeaxanthin* (Phe *z*)

73. *Phaeo-zeaxanthin* (Phe *z*)

74. *Phaeo-zeaxanthin* (Phe *z*)

75. *Phaeo-zeaxanthin* (Phe *z*)

76. *Phaeo-zeaxanthin* (Phe *z*)

77. *Phaeo-zeaxanthin* (Phe *z*)

78. *Phaeo-zeaxanthin* (Phe *z*)

79. *Phaeo-zeaxanthin* (Phe *z*)

80. *Phaeo-zeaxanthin* (Phe *z*)

81. *Phaeo-zeaxanthin* (Phe *z*)

82. *Phaeo-zeaxanthin* (Phe *z*)

83. *Phaeo-zeaxanthin* (Phe *z*)

84. *Phaeo-zeaxanthin* (Phe *z*)

85. *Phaeo-zeaxanthin* (Phe *z*)

86. *Phaeo-zeaxanthin* (Phe *z*)

87. *Phaeo-zeaxanthin* (Phe *z*)

88. *Phaeo-zeaxanthin* (Phe *z*)

89. *Phaeo-zeaxanthin* (Phe *z*)

90. *Phaeo-zeaxanthin* (Phe *z*)

91. *Phaeo-zeaxanthin* (Phe *z*)

92. *Phaeo-zeaxanthin* (Phe *z*)

93. *Phaeo-zeaxanthin* (Phe *z*)

94. *Phaeo-zeaxanthin* (Phe *z*)

95. *Phaeo-zeaxanthin* (Phe *z*)

96. *Phaeo-zeaxanthin* (Phe *z*)

97. *Phaeo-zeaxanthin* (Phe *z*)

98. *Phaeo-zeaxanthin* (Phe *z*)

99. *Phaeo-zeaxanthin* (Phe *z*)

100. *Phaeo-zeaxanthin* (Phe *z*)



Db 1231 CAGAGCAGCGCAAGATGCCAAGTCGCGCAAGAGACGCTTCATCGGATCGATGCTG 1290  
 Qy 244 ThrIleThrProSerLeuValAspGluLeuValAspAsnProAspPheProGluVal 263  
 Db 1231 ACCGTCAGCGCGGCTTATGAGGAGCGAGCAGCAGCATCGACTTCTTCATCATC 1350  
 Qy 264 SerSerGlyIleValValGlnGluValAlaProAsnSerProSerGluArgGlyVal 283  
 Db 1351 AGCGCGGGGATTTAGTCGACCGAGTGGTGGCCCTGTGTGCGCGCAGATTAAGGGGCAATC 1410  
 Qy 284 GlnAspGlyAspIleIleValIleValAsnGlyArgProLeuValAspSerSerGluLeu 303  
 Db 1411 AAAGAGGCGAGCTCTGCTGTCAGCTGAACGGCGC-CCTGTGGCAGCAGCGCGACCTG 1469  
 Qy 304 GlnGluValValLeuThrGluSerProLeuLeuGluValArgArgGlyAsnAspAsp 323  
 Db 1470 CAGGCGGCGCTGCGAGGAGGCGCGCTGCTGTGCGAGTGCGCAGGGGCAACGACGAC 1529  
 Qy 324 LeuLeuPheSerIleAlaProGluValValMet 334  
 Db 1530 CTGCTCTTCAACATCCAGCGGACCTCATCTCTG 1562  
 RESULT 5  
 BQ880316 890 bp mRNA linear EST 16-AUG-2002  
 LOCUS AGENCOURT\_8058818 NIH\_MGC\_102 Homo sapiens cDNA IMAGE:6213159  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BQ880316  
 VERSION BQ880316.1 GI:22272324  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 890)  
 NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-rcmail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: L10CM2378 row: 9 column: 16  
 High quality sequence stop: 578.  
 Location/Qualifiers  
 1. 890  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6213159"  
 /tissue\_type="epidermoid carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NIH\_MGC\_102"  
 /note="Organ: salivary gland; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Library constructed  
 by Ling Hong in the laboratory of Gerald M. Rubin  
 (University of California, Berkeley) using ZAP-cDNA  
 synthesis kit (Stratagene) and Superscript II RT (Life  
 Technologies). Note: this is a NIH\_MGC library."

Best Local Similarity: 91.10% Mismatches: 9  
 Query Match: 70.94% Indels: 12  
 DB: 5 Gaps: 2  
 US-10-617-443B-2 (1-334) x BQ880316 (1-890)  
 Qy 16 SerProArgTyrIleValPheAsnPheIleAlaAspValValGluValIleAlaProAlaVal 35  
 Db 2 AGCGCGGCTTCAAGATTCATCTTCATGCTGAGCTGAGAGATCGACCGCGTG 61  
 Qy 36 ValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSerSer 55  
 Db 62 GTCCACATAGAGCTTCTCTGAGACACCCGCTGTTGGCGCAACGTGCCCTCTCCAGC 121  
 Qy 56 GlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThrAsnAlaHisValValSer 75  
 Db 122 GGTTCGCTTCATCATGTCAAGAGCGGCTGTATCTATCAATCAATGCCAGTGGTGTC 181  
 Qy 76 SerAsnSerAlaAlaProGlyArgGlnGlnLeuValValGlnLeuGlnAsnGlyAspSer 95  
 Db 182 AGCAACAGTCTGCGCGCGGCGAGCAGCAGCTCAAGTGCATCAAGATGGGACTCC 241  
 Qy 96 TyrGluAlaThrIleValAspIleAspValAspValSerAspIleAlaThrIleValHis 115  
 Db 242 TATGAGGCGCACATCAAGACATCGACAGAGAGTCGAGCATTCGACCATCAAGATCCAT 301  
 Qy 116 ProValValValValValValValValValValValValValValValValValValVal 135  
 Db 302 CCCAAGAAAAAGCTCCCTGTGTGTTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361  
 Qy 136 PheValValAlaIleGlySerProPheAlaLeuGlnAsnThrValThrThrGlyIleVal 155  
 Db 362 TTGTGTGCGCATCGCAGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421  
 Qy 156 SerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyrIle 175  
 Db 422 AGCACTGCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 481  
 Qy 176 GlnThrAspAlaIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 195  
 Db 482 CAGCGGATGCCATCATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 541  
 Qy 196 GluValIleGlyIleAsnThrIleValValThrAlaGlyIleSerPheAlaIleProSer 215  
 Db 542 GAGGTCACTTGGCATCAACAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 601  
 Qy 216 Asp---ArgIleThrArgPheLeuThrGluPheGlnAspValGlnIleValAspTyr 234  
 Db 602 GAACCGATCAACACGGGTTTCTTCACAGAGTCCAAAGACAGCATCAAGACTGGGAA 661  
 Qy 234 GlyValArg-PheIleGlyIleArgMetArgThrIleThr-ProSer-LeuValAspGluLe 253  
 Db 662 GAAGCGGCTTCATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGAT 721  
 Qy 253 ValValAspSer-AsnProAspPheProGluVal---SerSerGlyIleIleThrVal-GlnGlu 272  
 Db 722 GAAGCGGAGCAACCGGAGCTTCCAGAGGTTTAAAGCTGGAATTTATGTGGCAAGAGG 781  
 Qy 272 AlaAla-ProAsnSerProSerGlnArgGly-GlyIleGlnAspGlyAsp---IleIleVal 290  
 Db 782 TTGCGCCCATTCATCTTCTTCAAGAGCGCGGATCAAAATGTGAACATCATCTGTC 841  
 Qy 290 IlyValValAsnGlyValArg 295  
 Db 842 AAGGTCACACGGCGG 857  
 RESULT 6  
 LOCUS CD635467 724 bp mRNA linear EST 12-JAN-2004  
 DEFINITION 56056242H1 FLP Homo sapiens cDNA, mRNA sequence.  
 ACCESSION CD635467  
 VERSION CD635467.1 GI:40283734  
 KEYWORDS EST.

ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.39e-133 Length: 890  
 Score: 1191.00 Matches: 266  
 Percent Similarity: 92.81% Conservative: 5

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 724)  
 Fu, G.K., Wang, J.T., Yang, J., Au-Young, J. and Stuve, L.L.  
 Circular rapid amplification of cDNA ends for high-throughput  
 extension cloning of partial genes  
 Genomics 84 (1), 205-210 (2004)  
 JOURNAL  
 COMMENT  
 Incyte Genomics, Inc.  
 3160 Porter Dr., Palo Alto, CA 94304, USA  
 Tel: 6508454102  
 Email: gfu@incyte.com.  
 FEATURES  
 source location/Qualifiers  
 1..724  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_lib="FLP"  
 /note="Vector: pDrive Cloning Vector"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6,76e-132 Length: 724  
 Score: 1176.00 Matches: 237  
 Percent Similarity: 98.35% Conservative: 1  
 Best Local Similarity: 97.93% Mismatches: 3  
 Query Match: 70.04% Indels: 0  
 DB: Gaps: 0  
 US-10-617-443B-2 (1-334) x CD635467 (1-724)  
 QY 80 A|A|P|G|I|Y|A|G|I|U|L|E|U|G|I|N|L|E|U|Y|S|E|A|P|S|E|T|Y|G|I|U|A|T|H| 99  
 D| 1 G|C|T|G|A|G|G|C|A|G|C|G|C|T|C|A|G|A|G|T|G|C|A|G|A|G|G|G|A|C|T|C|T|A|T|G|A|G|C|C|A|C| 60  
 QY 99 r|l|e|l|y|s|a|p|i|l|e|a|p|i|y|s|e|r|a|p|i|l|e|a|t|r|i|l|e|y|i|l|h|i|s|p|r|o|l|y|s|i|y| 119  
 D| 61 C|A|T|C|A|A|G|A|T|G|A|C|A|G|A|G|T|G|G|A|C|A|T|T|G|C|C|A|C|A|T|C|A|T|C|C|A|G|A|A|A| 120  
 QY 119 s|l|e|u|p|r|o|v|a|i|l|e|u|l|e|u|g|i|y|i|s|e|r|a|a|a|p|l|e|u|a|g|r|p|r|o|g|i|y|g|i|u|p|h|e|v|i|a|i| 139  
 D| 121 G|C|T|C|C|T|G|T|T|G|T|G|T|G|G|T|G|C|A|C|T|G|G|C|G|C|A|C|T|G|G|G|G|A|G|T|T|G|G|T|G|G|C| 180  
 QY 139 a|i|l|e|g|i|s|e|r|p|r|o|h|a|i|l|e|u|g|i|n|a|n|h|r|v|i|l|t|r|h|i|y|i|l|e|v|i|s|e|r|t|h|r|i|a|g|i| 159  
 D| 181 C|A|T|G|G|C|A|G|T|C|C|T|G|C|C|T|C|A|G|A|C|A|G|T|G|A|C|A|G|G|G|C|A|T|G|C|A|G|C|A|C|T|G|C|C|A| 240  
 QY 159 n|a|r|g|i|u|g|i|y|a|r|g|i|u|l|e|u|g|i|y|l|e|u|a|r|g|a|p|s|e|r|a|p|s|e|r|a|p|s|e|r|a|p|s|e|r|a|p|s|e|r| 179  
 D| 241 G|G|G|G|A|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G| 300  
 QY 179 a|i|l|e|l|e|a|n|h|r|y|g|i|y|a|n|s|e|r|i|g|i|y|p|r|o|l|e|v|i|a|n|l|e|u|a|p|i|y|g|i|u|v|i|l|e|g|i| 199  
 D| 301 C|A|T|C|A|C|A|C|T|A|G|G|G|A|A|C|C|C|G|G|G|G|A|C|C|A|C|G|T|G|A|A|C|T|G|A|T|G|G|G|G|G|G|G|G|G|G| 360  
 QY 199 y|i|l|e|a|n|h|r|l|e|u|v|i|a|i|t|h|r|i|a|g|i|y|i|s|e|r|p|h|e|i|a|i|l|e|p|r|o|s|e|r|a|p|i|g|i|l|e|h| 219  
 D| 361 C|A|T|C|A|C|A|G|G|C|T|A|A|G|G|T|C|A|G|G|G|C|G|C|A|C|T|C|T|T|G|C|A|T|C|C|C|T|C|A|G|C|G|G|A|C| 420  
 QY 219 r|a|r|g|h|e|l|e|u|h|t|h|g|i|u|p|h|e|i|n|a|p|i|y|s|e|i|n|i|l|e|y|s|a|p|i|t|r|p|i|y|l|e|y|a|r|p|h|i|l|e|g|i| 239  
 D| 421 A|C|G|G|T|C|C|T|C|A|G|A|G|T|T|C|C|A|G|A|C|A|G|C|A|G|C|A|A|A|G|C|T|G|A|A|A|C|G|C|T|T|C|A|C| 480  
 QY 239 y|i|l|e|a|r|g|e|t|a|r|g|h|i|t|h|r|p|r|o|s|e|r|i|u|v|i|a|a|p|i|y|l|e|u|l|e|u|y|i|l|e|s|e|r|a|n|p|r|o|s| 259  
 D| 481 C|A|T|C|A|C|A|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G| 540  
 QY 259 p|h|e|p|r|o|g|i|u|v|i|a|i|s|e|r|s|i|g|i|y|i|l|e|r|y|a|i|g|i|n|i|u|v|i|a|i|a|p|r|o|a|n|h|r|p|r|o|s|e|r|i| 279  
 D| 541 C|T|T|C|C|A|G|A|G|G|T|C|A|G|G|T|G|A|A|T|T|A|G|T|G|C|A|G|A|G|G|T|T|G|G|C|G|C|A|T|T|C|A|C|C|T|T|C| 600

QY 279 n|a|r|g|i|y|g|i|l|e|i|n|a|p|i|l|e|l|e|v|i|a|i|y|l|e|i|n|g|i|y|a|r|g|p|r|o|l|e|u|v|i|a|i|s| 299  
 D| 601 G|A|G|G|G|G|G|C|A|T|C|A|G|A|G|T|G|G|A|C|A|T|C|T|C|A|A|G|G|T|C|A|C|G|G|C|C|T|C|T|A|G|T|G|A| 660  
 QY 299 p|s|e|r|s|i|g|i|u|e|i|n|g|i|u|i|a|i|v|i|a|i|l|e|u|h|t|h|g|i|u|s|e|r|p|r|o|l|e|u|l|e|u|g|i|u|v|i|a|i|g|a|r| 319  
 D| 661 C|T|C|A|G|T|G|A|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G|G| 720  
 QY 319 g|e|i|y| 320  
 D| 721 G|G|G|A| 724  
 RESULT 7  
 CN787405  
 LOCUS  
 DEFINITION  
 CN787405 757 bp mRNA linear EST 26-MAY-2004  
 4121511 BARC 8BOV Bos taurus cDNA clone 8BOV\_34P04 5', mRNA  
 sequence.  
 ACCESSION  
 CN787405  
 VERSION  
 CN787405.1 GI:47683385  
 EST  
 KEYWORDS  
 Bos taurus (cow)  
 SOURCE  
 ORGANISM  
 Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.  
 1 (bases 1 to 757)  
 Baumann, R.G., Baldwin, R.L., Sonstegard, T.S., Van Tassell, C.P. and  
 Matukumalli, L.K.  
 Construction and Analysis of a cDNA Library Generated From  
 Intestinal Muscle and Epithelial Tissues of Holstein Cattle  
 Unpublished (2004)  
 Contact: Richard G. Baumann  
 ANRI  
 Bovine Functional Genomics Lab  
 BLDG 162: BARC-EAST, Beltsville, MD 20705, USA  
 Tel: 3015048604  
 Fax: 3015048744  
 Email: rbaumann@nri.barc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred  
 0.00925 using options -trim\_alt - -trim\_fastavector identified by  
 cross\_match using options -mismatch 12 -minscore 12  
 Plate: 34 row: P column: 04  
 Seq primer: CCTATTGATGACATCATATGAC  
 High quality sequence stop: 757.  
 FEATURES  
 source location/Qualifiers  
 1..757  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /strain="Holstein"  
 /db\_xref="taxon:9913"  
 /clone="8BOV\_34P04"  
 /sex="Female"  
 /tissue\_type="Epithelial, Muscle"  
 /dev\_stage="Lactating, Neonatal"  
 /lab\_host="DH10B TONa"  
 /clone\_lib="BARC 8BOV"  
 /note="Organ: Intestine; Vector: pCMVSPORT6.1; Site\_1:  
 library in pCMVSPORT6.1, constructed from equimolar mRNA  
 pools derived from 5 sources, 4 lactating intestinal, 1  
 neonatal intestinal 4/5 lactating, proximal duodenum,  
 jejunum, distal ileum, colon, 1/5 Neonatal, proximal  
 duodenum, jejunum, distal ileum"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.48e-129 Length: 757  
 Score: 1157.00 Matches: 233  
 Percent Similarity: 94.84% Conservative: 6  
 Best Local Similarity: 92.46% Mismatches: 13  
 Query Match: 68.91% Indels: 1  
 DB: Gaps: 0



Db	1177	TCCTTTGCCATCCCTCAGACCGCATCAGACGGTTCTCCTCAGAGTTCCAAAGCAACAG	1233
Qy	230		
Db	1237	ATCATA	1242
RESULT 9			
LOCUS	CO0404217		
DEFINITION	AGNCOCURT 27851724 NIH_MGC_252 Rattus norvegicus cDNA clone	796 bp	EST 01-JUL-2004
ACCESSION	CO0404217		
VERSION	CO0404217.1	GI:49586133	
KEYWORDS	EST.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
REFERENCE	1 (bases 1 to 796)		
AUTHORS	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics / NIH National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: <a href="mailto:cgapbbs-remail.nih.gov">cgapbbs-remail.nih.gov</a> Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical College of Wisconsin cDNA Library Preparation: Express Genomics cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLM15347 row: 0 column: 24 High quality sequence atop: 701. Location/Qualifiers		
FEATURES			
source	1. 796 /organism="Rattus norvegicus" /mol_type="mRNA" /db_xref="taxon:10116" /clone="IMAGE:7307090" /tissue_type="ovary, strain - Norway Line3. Agee weeks. Tissue was snap-frozen adn transferred in -70C. RNase free for the entire procedure" /lab_host="DH10B Tona" /clone_1ib="NIH MGC_252" /note="Organ: ovary; Vector: pDONR 201; Site_1: NoCl; Site_2: NoCl; RNA obtained from female ovariex animals at 8 wk old. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (TRI-reagent method). cDNA was primed using oligo-dT primer: 5'-pGACTGTTCTGATCGAGCGGCCCTCT-25-3' and cloned into the EcoRV/NotI sites of pExpress-1. Size-selection >1.25kb resulted in an average insert size of 1.7kb. This primary library is not normalized (normalized library is NIH_MGC_252) and was constructed by Express Genomics (Frederick, MD). Note: this is a NIH_MGC library"		
ORIGIN			
Alignment Scores:			
Pred. No.:	2,99e-119	Length:	796
Score:	1072.50	Matches:	216
Percent Similarity:	95.40%	Conservative:	12
Best Local Similarity:	90.38%	Mismatches:	10
Query Match:	63.88%	Indels:	2
Db:	7	Gaps:	1
US-10-617-443B-2 (1-334) x	CO0404217 (1-796)		

Oy		22	AasmPhell1eAlAaspValValGluylrlel1eAlAProAlVaValHlsrllegluePhe	41
Db		3	CACCTTATCCCGCATGTGGTGGAAGAATGCCCGCCAGCGGTGTTCATTATAGAGCTTT	62
Oy		42	LeuhArghIapProleuPhegiYARgaAnvaJProleuSerserglySerGIypheIIemet	61
Db		63	CTGAGGCACCCTGTGTTGGCCGAATGTGCCACTGTCCAGTGGCTCGGGCTTCATCATA	122
Oy		62	SerGIuaIaglyleuIIel1ethrraenAlAhIsvalVasereXpanseral1aAlapro	81
Db		123	TCAgAACCCGCTTGATCGTCACCAACGCCACAGTGTCTTCAGCTTCACAACCGCTTCA	182
Oy		82	GlyARgInglInleuLySVaIgInleuGlnaengIYApsertyrgIualathrIIelys	101
Db		183	GGTCGACGACGCTGAAGTGCACACTGCACAAATGCGGAGTCCATTATGAGGCCCACATCAG	242
Oy		102	AspIIeApLyalySserApII1ealAthrIIelysIIehIapOlyabLySvleuPro	121
Db		243	GACATCCACAAGAAGTCAGACATCGCCACCATCTTATTCACCCCACAACAAAAGCTCCC	302
Oy		122	VallenuleuengIyHIserralAapLeuAryProglyGluPheValAlaIalegly	141
Db		303	GTGTACTGCTGGGTCACTGGACAGACTCGGCTGGCCAGTITTGIGTGGCATTGGC	362
Oy		142	SerProPheAlaleuGlnAasnThralThrThrGIyll1evalserThralaglInaGlu	161
Db		363	AGTCCCTTTGGCCCTGCAAAACACCGGTACAAACGGGCACTGTACAGCACTGCCACGGGGAT	422
Oy		162	GlyARgIuLeuGlyLeuAryaapsSerApmMetApyrIIegInThraSPAlaIIele	181
Db		423	GGCAAAAGCTGGGCTCCGGGATTCGACATGGACATAHTCCAGACAGACGCATATC	482
Oy		182	AasnYrgIyaenserGIyGIYProleuValAasnLeuApsGIyuValIIegIyIAsn	201
Db		483	AAATRTGGGAACTCAGAGAGGCCCTCGTGTAACCTGATGATGATGATGCATCGCATCAAC	542
Oy		202	ThreuleySVaIIThralagIyIIeserPheAlAleprosearapaagIIethArxPhe	221
Db		543	ACGCTCAAAGGTGGGGCTGGTATCTCTTGGCCATCCCCCTCGAGTGCATCACCGGCTTC	602
Oy		222	LeuThrgIuPhegInAapLyGlnIIelyeAspTrpLySvArGPheIIegIyIAry	241
Db		603	CTCTGTAGTTCCAAAGACACGACATGTGAAGACCTGGAAGAAGCGCTTCATTGGCATCCGG	662
Oy		242	MeArghThrlIethrProSerLeuValApsGIuLeuLyalaSerrAenProApsPhe	260
Db		663	ATGCGGACATC-ACGCCAAAGTTGTGTGAGGACCTGAGACTGC--AAACCCAGATTTC	715
RESULT 10				
CV106197		798 bp	mRNA	linear EST 30-AUG-2004
LOCUS				
DEFINITION		AGENCOU0RT_31536328 NIH_MGC_269 Rattus norvegicus cDNA clone		
IMAGE:		7460898 5' , mRNA sequence.		
CV106197				
ACCESSION		CV106197.1 GI:51614770		
VERSION		EST		
KEYWORDS		Rattus norvegicus (Norway rat)		
SOURCE		Rattus norvegicus		
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
		Rattus.		
REFERENCE		1 (bases 1 to 798)		
AUTHORS		NIH-MGC http://mgc.nci.nih.gov/.		
TITLE		National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL		Unpublished (1999)		
COMMENT		Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics / NIH National Cancer Institute / NIH Bldg. 31 Rm10K07 Bethesda, MD 20892 Email: Gcgsbbs-remail.nih.gov Tissue Procurement: Dr. Josef Lazar, Dr. Howard J. Jacob, Medical College of Wisconsin cDNA Library Preparation: Open Biosystems		

CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

## FEATURES

## Source

```

/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="caxon:10116"
/clone="IMAGC:7460898"
/tissue_type="whole placenta, 2 pooled"
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_269"
/note="Organ: placenta; Vector: pExpress-1; site_1: EcoRV
site_2: NotI; Tissue was collected from two pooled
placentas from the 21st day of pregnancy. 1st strand cDNA
was primed with a Not I - oligo(dT) primer.
double-stranded cDNA was cloned into the Not I and EcoRV
sites of pExpress-1. Library was size-selected for >1.25
kb fragments for an average insert size of 2.05 kb. A
normalized version of this library is also available
(NIH_MGC 210). Library was constructed by Open Biosystems
(Huntsville, AL). Note: this is a Mammalian Gene
collection library"

```

## ORIGIN

**Alignment Scores:**

Pred. No.:	7.5e-118	Length:	798
Score:	1061.00	Matches:	219
Percent Similarity:	91.12%	Conservative:	17
Best Local Similarity:	84.56%	Mismatches:	20
Query Match:	63.19%	Indels:	6
DB:	7	Gaps:	2

US-10-617-443B-2 (1-334) X CV106197 (1-798)

QY	40	LeuPheLeuArgHisPLeuPheNegiYArGaaValPLeuSerSerGlySerGlyPhe	59
Db	32	ATCTTTCTGAAGCAACCTCTGTTTGGCCAAATGTCCACTGTCCAGTGGCTCGGGCTTC	91
QY	60	IleMetSerGluIaIaGlyLeuIleIerThraaMaIaHisValValSerSerAsnSerIa	79
Db	92	ATCATGTCAAGAAAGCCGGCTTGAATGTGCACAAAGCCACGTGTCTCCAGCTCCAAACCC	15
QY	80	AlaPArgGlyArGArgInIleuLybValGInLeuGInaenGlyAaPSerTyrGluIaThr	99
Db	152	GCCTCAGGTCTGGCAGCAGCTGAAGGTGCAACTGCAGAAATGGGGAGTCTTAAGAGCCACC	211
QY	100	IleAlaAspIleAspIyValSerAspIleAlaThrIleLysIleHisProLybLys	111
Db	212	ATCAGAGACATGCACAAAGTAGACATGCACACATCTTAATCACCCTCAACAAAG	277
QY	120	LeuProValLeuLeuLeuGlnIyHisSerIaAaPLeuArGProGlyGluPheValIaIa	135
Db	272	CTCCCGGTGTACTGTGTGGTCACTCGGAGACCTGCGGCTCGGGAGTTGTGTGGCC	333
QY	140	IleGlySerProPheAlaLeuGlnaenThrValThrThrGlyIleValSerThrIaGln	155
Db	332	ATCGGAGAGTCCCTTTTGCCTCCGCAAAACCTGTGACAAAGGGCATGTGACAGCATGCCAG	393
QY	160	ArgGluGlyArArgIleuLeuGlyLeuArGAsPAsrAsPMeIaSerTyrIleGlnThrAspAla	179
Db	392	CGGAGATGGCAAAAGCTGGGCTCTCGGGATTCGGACATGAGACTATATCAACAGAGCC	453
QY	180	IleIleAsnTyrGlyAaAsnSerGlyGlyPLeuLeuValaenLeuAspGlyGluValIleGly	199
Db	452	ATCATCAATTATGCGGAACTCAGAGGCCCCCTGGTGAACCTGGAGTGGTGAAGTATCGGC	511
QY	200	IleAsnThrLeuLybValThrIaGlyIleSerPheAlaIleProSerAspArgIleThr	215

Db	512	ATCAACAGCGCTCAAGGTGGCGGCTGGTATCTCTTCCTTGGCATCCCTCGGATCGCATCAAG	571
Qy	220	ArgPheLeuThrGluPheGlnAspIysGlnIleIysAspTrpLysLeuArgPheIleGly	239
Db	572	CGCTTCTCTCTCGAGTTCCAGACACAGCATGGAAAGACTGMAAGAGGCTTCATTGAC	631
Qy	240	IleArgMetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAsnProAsp	259
Db	632	ATCCGGAATCCGACCATCATCCCAAGTTGGTGGAGGAGAACTGMAACTGCCAACCCAGAT	691
Qy	260	PheProGluValSerSerGlyIleIleTyValGlnGluValAlaIleProAsnSerProSerGln	279
Db	692	TTTCA-GCAGTGAGCAGTGATATATGTTCA-GAGGTGGTCCACTCATCTTCAGAG	748
Qy	280	ArgGlyGlyIleGlnAspGlyAspIleIleValIysValAsnGlyValArgProLeuVal	298
Db	749	AGA---GGCATCTCGATGCGA-----CTCATATCAAGCTCATGAGCTCCCTTAGTT	796

## RESULT 11

AY40604 /  
100710

## DEFINITION

genomic survey sequence.

## KEYWORDS

VERSION AY406047.1 GI:39762021

**DOOR TO DOOR ORGANIS**

## AUTHORS

Adams, M.D. and Cargill, M.

## TOTENAT

JOURNAL OF CLIMATE 302 (5652), 1960-1963 (2003)  
PUBMED 14571303

## REFERENCE AUTHORS

ACIMONS  
CIAZIN, R. O.; GIANOWSKI, S.; WIEBULL, R.; INOUBI, F.; KEGALIM, A.;  
TODD, M. A.; TANENBAUM, D. M.; CIVELLO, D. B.; IM, E.; MURPHY, B.

Adams, M.D. and Cargill, M.

**JOURNAL**

Rockville, MD 20850, USA

**SECRET**

FEATURES	Location/Qualifiers
Source	1 073

name

```
gene
  <1: 2513
  /gene="PRSS11"
```

## ORIGIN

## Alignment

**Score:**

Pred. No.:	1.69e-114	length:	97
Score:	1034.50	Matches:	2032
Percent Similarity:	80.43%	Conservative:	60
Best Local Similarity:	62.08%	Mismatches:	55
Query Match:	61.61%	Indels:	9
DB:	9	Gaps:	2

US-10-617-443B-2 (1-334) X AYY406047 (1-973)

Oy 10 GlyLeuHicIleuSerSeProlGlyrLysepAsnPhoileAlaaspValValGlu 29  
::|||::|||::|||::|||::|||  
2 GGGCAGGAAAGATCCCAACAGTTTGGCCCATTAATATATTACTTTATCGGCAGCTGTGGAG 61

```

QY 30 LysIleAlaProAlaValAlaHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
Db 62 AAGATGCCCCCTGCGGTTCATATGCAATGTTTCGCAAGCTTCGTTTTCTTAAACCA 121
QY 50 AannValProLeuSerSerGlySerGlyPheIleMetSerGlnAlaGlyLeuIleIleThr 69
Db 122 GAGGTGCCCCGTCGTGGTGTCTGGTTATGTTCTCGAAGATGACATGATCGAGACA 181
QY 70 AannAlaHisValAlaValSerSerAenSerAlaAlaProGlyArgGlnGlnLeuValGln 89
Db 182 AATGCCACGTGTGTGCAAC-----AAGACCGGGTCAAGTTGAG 223
QY 90 LeuGlnAenGlyAenSerGlyGlyAlaThrIleLeuAspIleLeuAspIleValSerAspIle 109
Db 224 CTGAAGAAGCGTGCCCTTAACGAAGCCAAATCAAGATGTGATGAGAAGACGACATC 283
QY 110 AlaThrIleLeuSerIleHisProGlySerGlyLeuProValLeuLeuGlnGlyHisSerAla 129
Db 284 GCATCATCAAAATTAATGACCAACGGGCAAGCTCTCTCTGCTGCTGGCCCTCTCA 343
QY 130 AspLeuArgProGlyGluPheValAlaAlaIleGlySerProPheAlaLeuGlnAenThr 149
Db 344 GAGCTCGCGCGCGGAGGTTTCGTGTCGTCATCGAAGCCCGTTTCCCTTCAAAACACA 403
QY 150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAsp 169
Db 404 GTCAACACCGCGGATCGTAGACACACCCGAGCGGCAAGACGCGGCGCTGCGCCAC 463
QY 170 SerAspMetAspIleIleGlnThrAspAlaIleIleAsnIleGlyAsnSerGlyGlyPro 189
Db 464 TCAGACATGACTACATCCAGACCGACCATCATCACTATGAAACTCGGAGGCGCG 523
QY 190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrIleuValThrAlaGlyIle 209
Db 524 TTAGTAAACCTGACCGGTGAAGTGAATTAACATTGAAAGTGAACGCTGGAATC 583
QY 210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspIleVal 229
Db 584 TCCCTTGCAATCCCATCTGATGAATTAAGTTCTCTCGAGATGCCATGACCGACAG 643
QY 230 IleLeuAsp-----TrpIleValArgPheIleGlyIleArgMetArgThrIleThr 246
Db 644 GCCAAAGAAAGCCATCACCAGGAAGATATATGTTATCCGAATGATGATCATCAG 703
QY 247 ProSerLeuValAspGluLeuValAsnSerAspProAspPheProGluValIleSerGly 266
Db 704 TCCAGCAAAAGCAAGAGCTGAAGGACCGGACCGGACCTCCAGACGTGATCTCAGGA 763
QY 267 IleThrValGlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGly 286
Db 764 GCGATATATATTAAGTAATTCCTGATATCCCGACCAAGCTGTGTCTCAAGAAAC 823
QY 287 AspIleIleValIleValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAla 306
Db 824 GAGCTCATATATGACATCAATGACATGACATGCGTGTCTCCGCAATGATGTCAGCGAGTC 883
QY 307 ValLeuThrGluSerProLeuLeuGluValAlaArgArgIleAsnAspAspLeuPhe 326
Db 884 ATTAAGAAAGGAAAGCACCTGAACATGTGTGTCCGAGGGGTAAATGAATATCATGATC 943
QY 327 SerIleAlaProGluValAla 333
Db 944 ACGATGATTCGCAAGAAATT 964

```

RESULT 12  
 CRS90731 1655 bp mRNA linear HTC 21-JUN-2004  
 LOCUS full-length cDNA clone CSDBE011YH20 of Placenta of Homo sapiens  
 DEFINITION (human)  
 ACCESSION CRS90731 GI:50471538  
 VERSION CRS90731.1  
 KEYWORDS HTC; CNSLT\_cDNA.

```

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1655)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@life.rockefeller.edu
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
REFERENCE 2 (bases 1 to 1655)
AUTHORS Faraday Avenue
TITLE Genoscope.
JOURNAL Direct Submission
SUBMITTED (20-JUN-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: sequef@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
source 1..1655
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSDBE011YH20"
/libase_type="Placenta"
/plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.: 3,836-114 Length: 1655
Score: 1034.50 Matches: 206
Percent Similarity: 80.18% Conservative: 61
Best Local Similarity: 61.86% Mismatches: 57
Query Match: 61.61% Indels: 10
DB: Gaps: 2
US-10-617-443B-2 (1-334) x CRS90731 (1-1655)
QY 4 AlaLeuProAlaSerAlaGlyLeuHisIleGluLeuSerProArgIleValPheAspIle 23
Db 136 GCGAGCCCTGCGGCGCAAGGCGAG-GAAGATCCCAACAGTTTGGCCATTAATTAATCTT 194
QY 24 IleAlaAspValAlaGlnValIleAlaProAlaValAlaHisIleGluLeuPheArg 43
Db 195 ATCGGCGAGGTGTGTGAAGATCCCGCTGCGGTTCATATGCAATGTTTGCAGAG 254
QY 44 HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGlu 63
Db 255 CTTCCTTTTCTTAACGAGAGGTGCGGCGGTAGTGGGTCTGGGTTATGTGTGCGAA 314
QY 64 AlaGlyLeuIleIleThrAsnAlaHisValAlaSerSerAsnSerAlaAlaProGlyArg 83
Db 315 GATGACATGATGTGTGCAAAATGCCACGCTGTGACCAAC-----AAG 356
QY 84 GlnGlnLeuValAlaGlnGlnAsnGlyAspSerGlyGluAlaThrIleLeuAspIle 103
Db 357 CACCGGTTCAAGTGTGAGCTGAAGAGCGGTGACATCTTAAGAAACCAAAATCAAGATG 416
QY 104 AspIleValSerAspIleAlaThrIleValIleHisProGlyValLeuProValLeu 123
Db 417 GATGAAAGACAGATCGACATCAAAATTAAGACCAACGAGCAAGCTGCTGCTCG 476
QY 124 LeuLeuGlyHisSerAlaAspLeuArgProGlyGluPheValAlaIleGlySerPro 143
Db 477 CTGCTTGCGCGCTCCCAAGCTGCGCGCGGAGATGTTGTGTGCTGCAATCGAAGCCG 536
QY 144 PheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArgGlyArg 163
Db 537 TTTTCTTCAAAACACATGATCACACCGGATCTGTAGACACCAACGAGGCGGCAAA 596

```



```

Qy 164 GlnleuGlyLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyr 183
Db 597 GAGCTGGGGGCTCCGCAATCAGACATGACATCCAGACCGCATCATCATAT 656
Qy 184 GlysAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeu 203
Db 657 GGAACATCGGAGAGCCCGTTAGTAACCTGAGCGGTGAAGTGAATTGAATTAACACTTGG 716
Qy 204 LysValThrIleGlyIleSerPheAlaIleProSerAspGlyIleThrArgPheLeuThr 223
Db 717 AAGGTGACAGTGGAAATCTCTTGGCAATCCCATGTAGATTAAGATAAAAGTTCTCAGC 776
Qy 224 GluPheGlnAspLysGlnIleLysAsp-----TrpLysLysArgPheIleGlyIle 240
Db 777 GAGTCCCATGACCAACAGAGCCAAAGAAAGCCATCACAGAAAGAAAGTATATTTGATC 836
Qy 241 ArgMetArgThrIleThrProSerLeuValAspGluLeuLysAsnProAspPhe 260
Db 837 CGAATGATGTCACTCAGTCAGCAAGCCAAAGAGTGAAGAGACCGGACCGGACTTC 896
Qy 261 ProGluValSerSerGlyIleTyrValGlnGluValAlaProAsnSerProSerGlnArg 280
Db 897 CCAAGACGTGATCTCAGAGCGGTATATTAATGAAGTAAATTCCTGATACCCAGCAAGACT 956
Qy 281 GlnGlyIleGlnAspGlyAspIleIleValLysValAsnGlyArgProLeuValAspSer 300
Db 957 GGTGGTCTCAAGAAACGAGCGTATATATGACATCAATGAGACGTCGGGTGTCCTCCGC 1016
Qy 301 SerGluLeuGlnGluValAlaValLeuThrGluSerProLeuLeuGluValArgArgGly 320
Db 1017 AATGATGTACGACGATCATTAATAAGGAAAGCAACCTGAAGACATGATGTCGAGGGGT 1076
Qy 321 AsnAspAspLeuLeuPheSerIleAlaProGluValVal 333
Db 1077 AATGAAGATATCATGATCAGACGATTCCTCGAAGAAATTT 1115

RESULT 13
BC011352 1911 bp mRNA linear HTC 19-NOV-2003
LOCUS Homo sapiens protease, serine, 11 (IGF binding), mRNA (cDNA clone
IMAGE:417882), containing frame-shift errors.
ACCESSION BC011352
VERSION BC011352.1 GI:15030191
KEYWORDS HTC
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
1 (bases 1 to 1911)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heile, F.,
Datchenko, L., Marusina, K., Farmer, A.F., Rubin, G.W., Hong, L.,
Shapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Tomihata, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abrams, R.D., Muliyil, S.J., Bosak, S.A., McEwan, P.J.,
Mokeman, K.J., Malek, J.A., Guarnatone, P.H., Richard, S.,
Worley, K.C., Hale, J.A., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Vallalun, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahney, J., Helton, E., Kettman, M., Madan, A., Kodrich, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Dickson, M.C., Rodriguez, A.C., Greenwood, J.W., Green, E.D.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.,
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL MEDLINE 22388257
PUBMED 12477932

```

```

REFERENCE 2 (bases 1 to 1911)
AUTHORS Straussberg, R.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genome Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Guarnatone, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louieged, H.,
Kowis, C.R., Speed, A.J., Martin, R.G., Muzny, D.M., Navarati,
A.N., Gibbs, R.A.
FEATURES
source
1. 1911
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:417882"
/tissue="Brain, anaplastic oligodendroglioma with
1p/19q loss"
/clone_id="NCI CGAP_Brn67"
/lab_host="DHIOB"
/note="Vector: pCMV-SPORT6"
ORIGIN
Alignment Scores:
Pred. No.: 4,78e-114 Length: 1911
Score: 1034.50 Matches: 206
Percent Similarity: 80.18% Conservative: 61
Best Local Similarity: 61.86% Mismatches: 57
Query Match: 61.61% Indels: 10
DB: 3 Gaps: 2
US-10-617-443B-2 (1-334) x BC011352 (1-1911)
Qy 4 AlaLeuProAlaSerAlaGlyLeuHisGlnLeuSerSerProArgTyrIlePheAsn 23
Db 350 GCGGACCTCGGGCCAAAGGGCAG-GAAGATCCCAACAGTTGGCCATTAATTAATCTTT 408
Qy 24 ILeAlaAspValValGluLysIleAlaProAlaValAlaHisIleGluLeuPheLeu 43
Db 409 ATCGCGAGCGTGTGAGAAAGTCGCCCTCGCGTGTTCATATCGAATGTGTTCCCAAG 468
Qy 44 HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGly 63
Db 469 CTTCCTGTTTCTTAACAGAGAGTCCCGGTGCTAGTGGGTCTGGGTTTATGTCGGA 528
Qy 64 AlaGlyLeuIleIleThrAsnAlaHisValAlaSerSerAsnSerAlaAlaProGlyArg 83
Db 529 GATGACCTAGCTGATCAATGCGCCAGCGTGACCAAC-----AAG 570
Qy 84 GlnGlnLeuLysValGlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIle 103
Db 571 CACCGGCTCAAGATGACGTGAAGACGCGTGCACCTTAAGAAACCAAAATCAAGATGCG 630
Qy 104 AspLysSerAspIleAlaThrIleLysIleHisProLysLysValLeuProValLeu 123

```

Db 631 GATGAGAAAGCAGACATCGCACTCATCAAAATTGACACAGGGGAGGCTGCTCTCTG 690  
 Qy 124 LeuLeuGlyHisSerAlaAspLeuArgProGlyGluPheValAlaIleGlySerPro 143  
 Db 691 CTGCTTGCGCGCTCTCTCAGAGCTCGCGCGGAGAGTTCTGTGTGCTCCATCGGAAGCCCG 750  
 Qy 144 PheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArgGluGlyArg 163  
 Db 751 TTTTCCCTTCAAAACAGTCACACCGAGGAGTGTGAGACACACCGAGGAGCGGCAAA 810  
 Qy 164 GluLeuGlyLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyr 183  
 Db 811 GAGCTGGGGTCTCGCAACTCAGACATGACATCAGACCGACCGCTCACTCACTAT 870  
 Qy 184 GlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeu 203  
 Db 871 GGAAACTCGGAGAGCCCGTTTAACTGAGACGAGTGAAGTGTGAAATTAACTTTG 930  
 Qy 204 LysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThr 223  
 Db 931 AAAGTGCACCTGGAATCTCTTTCGAAATCCATCGATTAAGATTAAAGTTCTCAGC 990  
 Qy 224 GluPheGlnAspLysGlnIleLysAsp-----TrpLysLysArgPheIleGlyIle 240  
 Db 991 GAGTCCCATGACCGACAGCCCAAGAAAGCCATCACCAAGAAAGATTATTTGATTC 1050  
 Qy 241 ArgMetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAspProAspPhe 260  
 Db 1051 CGAATGATGTCATCTCAGCTCCAGCAAAAGCCAAAGAGCTGAAGACCGGACCGGACTTC 1110  
 Qy 261 ProGluValSerSerGlyIleTyrValGlnGluValAlaProAsnSerProSerGlnArg 280  
 Db 1111 CCAACGCGAATCTCAGAGCGCTATATATGAAGTAATTCCTGATACCCGACAGAGAGCT 1170  
 Qy 281 GlyGlyIleGlnAspGlyAspIleIleValLysValAsnGlyArgProLeuValAspSer 300  
 Db 1171 GGTGTCTCAAGAAAGACAGCTCATATACGATCAATGACATGACGTCGCGGTCTCGCC 1230  
 Qy 301 SerGluLeuGlnGlnAlaValLeuThrGluSerProLeuLeuGluValAlaGlyArgGly 320  
 Db 1231 AATGATGTCACGACGATCATTAAGGAAAGCACCCGAAACATGATGTCGCGAGGGGT 1290  
 Qy 321 AsnAspAspLeuLeuPheSerIleAlaIleProGluValVal 333  
 Db 1291 AATGAAATATCATGATCATCAGTATTCCTCCGAAAGAAATT 1329  
 Db  
 RESULT 14  
 CR603877 1931 bp mRNA linear HTC 21-JUN-2004  
 LOCUS full-length cDNA clone CS0D1075YNO6 of Placenta Cot 25-normalized  
 DEFINITION  
 CR603877  
 VERSION CR603877.1 GI:50484684  
 KEYWORDS HTC; cNSLT cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 1931)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 CONTACT: Feng Liang Email: fliang@life.uci.edu URL: http://fulllength.invitrogen.com/ invitrogen Corporation 1600  
 REMARK Faraday Avenue  
 2 (bases 1 to 1931)  
 REFERENCE Genoscope.  
 AUTHORS Direct Submission  
 TITLE Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage :  
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
 COMMENT - Web : www.genoscope.cns.fr)  
 end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
 location/Qualifiers  
 1. 1931  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1075YNO6"  
 /issue\_type="Placenta Cot 25-normalized"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4.86e-114 length: 1931  
 Score: 1034.50 matches: 206  
 Percent Similarity: 80.18% conservative: 61  
 Best Local Similarity: 61.86% mismatches: 57  
 Query Match: 61.61% indels: 10  
 Gaps: 2  
 DB:  
 US-10-617-443b-2 (1-334) x CR603877 (1-1931)  
 Qy 4 AlaleuProAlaSerAlaGlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPhe 23  
 Db 454 GCGAGCGCTCGCGCCAGAGGCGAG-GAAGATCCCAACAGTTGGCGCCATTAATTAATTCTT 512  
 Qy 24 ILeAlaAspValAlaGluValIleAlaProAlaValAlaHisIleGluLeuPheLeuArg 43  
 Db 513 ATCGGAGCGTGGTGGAGAAATGCGCCCTGCGGTGTTCAATCGAATGTTTCCGACAG 572  
 Qy 44 HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGlu 63  
 Db 573 CTTCGCTTTCTTAAACGAGAGCTCGCGGTGCTGAGTGTATGTTGTCGAA 632  
 Qy 64 AlaGlyLeuIleIleThrAsnAlaHisValLysSerAsnSerAlaAlaProGlyArg 83  
 Db 633 GATGACCTGATCGTCAAAATGCCACGTGTGACCAAC-----AAG 674  
 Qy 84 GlnGlnLeuLysValGlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIle 103  
 Db 675 CACCGGCTCAAAAGTTGAGCTGAAGAAGCGTGCACCTTACGAAGCAAAATCAAGAGTGTG 734  
 Qy 104 AspLysLysSerAspIleAlaThrIleLysValHisProLysLysValLeuProValLeu 123  
 Db 735 GATGAGAAAGCAGACATCGCATCAAAATTGACACCGGACAGCTCTCTCTCTG 794  
 Qy 124 LeuLeuGlyHisSerAlaAspLeuArgProGlyGluPheValAlaIleGlySerPro 143  
 Db 795 CTGCTTGCGCGCTCTCTCAGAGCTCGCGCGGAGAGTTCTGTGTGCTCCATCGGAAGCCCG 854  
 Qy 144 PheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArgGluGlyArg 163  
 Db 855 TTTTCCCTTCAAAACACAGTCACACCGGAGTCGAGACACACCGAGCGGCGCAAA 914  
 Qy 164 GluLeuGlyLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyr 183  
 Db 915 GAGCTGGGGTCTCGCAACTCAGACATGACATCACTCAAGCCGACCATATCAACTAT 974  
 Qy 184 GlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeu 203  
 Db 975 GGAAACTCGGAGAGCCCGTTTAACTGAGACGAGTGAAGTATGAAATTAACTTTG 1034  
 Qy 204 LysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThr 223  
 Db 1035 AAAGTGCACCTGGAATCTCTTTCGAAATCCATCGATTAAGATTAAAGTTCTCAGC 1094  
 Qy 224 GluPheGlnAspLysGlnIleLysAsp-----TrpLysLysArgPheIleGlyIle 240  
 Db 1095 GAGTCCCATGACCGACAGCCCAAGAAAGCCATCACCAAGAAAGATTATTTGATTC 1154  
 Qy 241 ArgMetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAspProAspPhe 260





**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2005, 18:09:44 ; Search time 4078 Seconds  
(without alignments)  
9688.747 Million cell updates/sec

Title: US-10-617-443B-1\_COPY\_1\_1038

Perfect score: 1038  
Sequence: 1 caggagactcgagacttcgag.....ggccctccgcgcagcgag 1038

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 34239544 seqs, 19032134700 residues

Word size : 12

Total number of hits satisfying chosen parameters: 2836061

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : EST.\*

1: gb\_esc1.\*  
2: gb\_esc2.\*  
3: gb\_esc3.\*  
4: gb\_esc4.\*  
5: gb\_esc5.\*  
6: gb\_esc6.\*  
7: gb\_esc7.\*  
8: gb\_esc8.\*  
9: gb\_esc9.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	3.5	625	9	AG120009 Pan trogl
2	24	2.3	732	5	B0574834
3	24	2.3	2534	3	BSM804741
4	23	2.2	728	4	BG548488
5	22	2.1	125	4	BM481069
6	22	2.1	498	6	CD547336
7	22	2.1	505	6	CD554239
8	22	2.1	744	6	CB528031
9	22	2.1	922	2	BF233030
10	22	2.1	1706	5	BU409267
11	22	2.1	1993	3	AK079109
12	22	2.1	3268	3	AK040893
13	22	2.0	204	4	BI329065
14	22	2.0	207	7	CR516621
15	21	2.0	238	2	BB200500
16	21	2.0	243	5	BU756781
17	21	2.0	247	9	CC467496
18	21	2.0	283	1	AI840198
19	21	2.0	314	2	AMS21894
20	21	2.0	322	1	AI575250
21	21	2.0	322	1	AI705385
22	21	2.0	324	7	CR839023
23	21	2.0	327	2	AW212123
24	21	2.0	345	7	CP135804

25	21	2.0	359	4	BI301529	BI301529	UI-R-DNO-
26	21	2.0	375	4	BG381393	BG381393	UI-R-CSO-
27	21	2.0	388	6	CB794415	CB794415	AMGNMNC-U
28	21	2.0	397	2	BF417430	BF417430	UI-R-CNO-
29	21	2.0	405	5	BY367762	BY367762	BY367762
30	21	2.0	407	1	AA967928	AA967928	uh02h11.x
31	21	2.0	426	1	AI553578	AI553578	vw42c02.x
32	21	2.0	428	2	BE860202	BE860202	UI-R-AO0-
33	21	2.0	428	7	CN259060	CN259060	170006000
34	21	2.0	470	1	AI143088	AI143088	DKP2547F
35	21	2.0	508	1	AI415257	AI415257	md17f12.x
36	21	2.0	540	2	BE691054	BE691054	uv55b11.x
37	21	2.0	545	5	BP213364	BP213364	BP213364
38	21	2.0	582	5	BP288294	BP288294	BP288294
39	21	2.0	582	5	BP314507	BP314507	BP314507
40	21	2.0	582	5	BP314507	BP314507	BP314507
41	21	2.0	582	5	BP352419	BP352419	BP352419
42	21	2.0	603	5	BQ210450	BQ210450	UI-R-DY1-
43	21	2.0	604	6	CD371381	CD371381	UI-R-DY0-
44	21	2.0	614	7	CV029025	CV029025	7590_Full1
45	21	2.0	619	7	CN259057	CN259057	170004706

## ALIGNMENTS

RESULT 1  
LOCUS AG120009  
DEFINITION Pan troglodytes DNA, clone: PTB-128K01.F, genomic survey sequence.  
ACCESSION AG120009  
VERSION AG120009.1 GI:16649174  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
REFERENCE  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE BAC end sequences of library PTB  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 625)  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimbases@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.  
PRIMERS  
Sequencing: -21M13  
LIBRARY Vector : pKS145  
R-site 1 : SacI  
R-site 2 : SacI  
Location/Qualifiers  
1. 625  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="PTB-128K01.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
ORIGIN  
Query Match 3.5%; Score 36; DB 9; Length 625;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGGACTCGAAGTTGACTCTCCACACTGCTT 36  
 DB 590 CAGGACTCGAAGTTGACTCTCCACACTGCTT 625

RESULT 2  
 LOCUS B0574834/c

DEFINITION UI-H-EZ1-bbc-f-19-0-UI.s1 NCI\_CGAP\_Ch2 Homo sapiens cDNA clone

ACCESSION B0574834  
 VERSION B0574834  
 KEYWORDS EST

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Auld,C., Osanger,A.,  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strauberg, Ph.D.  
 Email: cgaaps-remail.nih.gov  
 Tissue Procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of  
 Orthopaedics  
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 176-463, >ALU (matched complement)  
 Seg primer: M13 FORWARD  
 POLYA=Yes.

FEATURES  
 source

1..732  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-EZ1-bbc-f-19-0-UI"  
 /tissue\_type="Chondrosarcoma Grade II"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI\_CGAP\_Ch2"  
 /note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)  
 with a modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 NCI CGAP Ch2 is a normalized cDNA library containing the  
 following tissue(s): Chondrosarcoma Grade II. The library  
 was constructed according to Bonaldi, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 TGATCAGCT.  
 TAG TISSUE=grade-2-chondrosarcoma  
 TAG\_LIB=UI-H-EZ1  
 TAG\_SEQ=ATCTAATATG"

ORIGIN

Query Match 2.3%; Score 24; DB 5; Length 732;  
 Best Local Similarity 100.0%; Pred. No. 1.7;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 TGAGGCCACACAGCTTGAAAGCAG 141  
 DB 493 TGAGGCCACACAGCTTGAAAGCAG 470

RESULT 3  
 LOCUS HSM604741

DEFINITION Homo sapiens mRNA; cDNA DKFZp313M0539 (from clone DKFZp313M0539).

ACCESSION AL833428  
 VERSION AL833428.1 GI:21734070

KEYWORDS HTC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Auld,C., Osanger,A.,  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strauberg, Ph.D.  
 Email: cgaaps-remail.nih.gov  
 Tissue Procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of  
 Orthopaedics  
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 176-463, >ALU (matched complement)  
 Seg primer: M13 FORWARD  
 POLYA=Yes.

FEATURES  
 source

1..2534  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="RZPD:DKFZp313M0539"  
 /db\_xref="taxon:9606"  
 /clone="DKFZp313M0539"  
 /tissue\_type="heart"  
 /clone\_lib="313 (synonym: h1cc2). Vector pT73-Pac; host  
 DH10B; sites SfiI + SfiIb"  
 /dev\_stage="adult"  
 /note="esophageal cancer associated protein, C-terminal  
 truncated, not fully spliced"  
 1..2534  
 /gene="DKFZp313M0539"  
 5..>1228  
 /gene="DKFZp313M0539"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="CAH10399.1"  
 /db\_xref="GI:50949538"  
 /translation="MAVPMHSMRNRNYKAFASCRLEAVPLRFGDYHPLKPTYESK  
 TKRNKSTSTSSSSSVYDPLSLDGTDPISMPATRADPAALAAANDSSRRK  
 DRDDSVVGSDEPWTNRGRIILAVYTTKLSINLEWGSKGAKATITLMSEKVT  
 RLEELDDSEBSQKELNLVTOQDVNRIEELNOSIKAMASDOKYKALITVICSKL  
 SDTSVIOYPSKFTVITIDTDFGLVNERFSCVDSRSVLPHPFSSENNDAKET  
 CLNMFKFSRIELIPREYVASTIKCKPKFISCLPRLTICMIGIDIPVSVY  
 ARAVTCRYGMAVAPLHKETLNKNPFDFLTTRKQIHGDIVQNVGVGLSPVLYP  
 PAMDTIFOCISTHAPF"

ORIGIN

Query Match 2.3%; Score 24; DB 3; Length 2534;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 TGAGGCCACACAGCTTGAAAGCAG 141  
 DB 1468 TGAGGCCACACAGCTTGAAAGCAG 1491

RESULT 4  
 LOCUS BG548488  
 728 bp mRNA linear EST 04-APR-2001

DEFINITION	602575077p1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4703322 5'
ACCESSION	BC548488
VERSION	BC548488.1 GI:13547153
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 728)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plates: LINC1541 row: 1 column: 19 High quality sequence stop: 695. location/Qualifiers
FEATURES	1..728
SOURCE	<pre> /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:4703322" /lab_host="DH10B (TI phage-resistant)" /clone_id="NIH_MGC_77" /note="Organ: Lung; Vector: pDNR-LIB (Clontech); site 1: SfiI (ggcgccgcgcgcgc); site 2: SfiI (ggcgccatcgcc); 5' adaptor sequence: 5'-CAGCGCATTTATGCGC-3' and 3' adaptor sequence 5'-ATTCTAGAGCGCCGAGCGCGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 clones contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC library." </pre>
ORIGIN	
Query Match	2.2%; Score 23; DB 4; Length 728;
Best Local Similarity	100.0%; Pred. No. 5.4;
Matches	23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	139 CAGCCAGGCCGAGCCGGAACCCCT 161
DB	491 CAGCCAGGCCGAGCCGGAACCCCT 513
RESULT 5	
LOCUS	BM481069 125 bp mRNA linear EST 05-FEB-2002
DEFINITION	532861 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION	BM481069
VERSION	BM481069.1 GI:18531397
KEYWORDS	EST.
SOURCE	Bos taurus (cow)
ORGANISM	Bos taurus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
AUTHORS	1 (bases 1 to 125) Smith,T.P.L., Grose,W.M., Feeking,B.A., Roberts,A.J., Stone,R.T., Casae,E., Wray,J.E., White,J., Cho,J., Fahrnkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckow,C.G., Pertea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

```

JOURNAL      Genome Res. 11 (4), 626-630 (2001)
MEDLINE      21180013
COMMENT      11282978
              Contact: Smith TPL
              USDA, ARS, US Meat Animal Research Center
              PO Box 166, Clay Center, NE 68933-0166, USA
              Tel: 402 762 4366
              Fax: 402 762 4390
              Email: smtth@mail.marc.usda.gov
              Single pass sequencing. Bases called and alt. trimmed with phred
              v0.980904.e. Vector identified by cross_match with the -minscore 18
              and -mismatch 12 options.
              PCR primers
              FORWARD: AGGAAACAGCTATGACCAT
              BACKWARD: GTTTCCTCAGTCACGACG
              Plate: 2      row: G      column: 20
              Seq primer: ATTTAGGTGACACTATGT.
              Location/Qualifiers
                1..125
                  /organism="Bos taurus"
                  /mol_type="mRNA"
                  /db_xref="taxon:9913"
                  /rlnase_type="pooled"
                  /lab_host="DH10B"
                  /clone_lib="MARC 4BOV"
                  /note="Vector: PCMV SPORT6; site_1: NotI; site_2: SalI;
                  library made from pooled tissue from day 20 and day 40
                  embryos."

ORIGIN
Query Match      2.1%; Score 22; DB 4; Length 125;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      329  GGTCCTGGCAGCCACAGCCCT 350
          |||||
          6  GGTCCTGGCAGCCACAGCCCT 27

RESULT 6
LOCUS      CD547336 498 bp mRNA linear EST 11-JUN-2003
DEFINITION      B0281B08-5 NIA Mouse Embryonic Germ Cell cDNA Library (long) Mus
ACCESSION      CD547336
VERSION        CD547336.1 GI:31595071
KEYWORDS       EST.
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 498)
Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
JOURNAL      MEDLINE
MEDLINE      21429098
COMMENT      11544199
              Contact: Dawood B. Dudekula
              Laboratory of Genetics
              National Institute on Aging/National Institutes of Health
              333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
              Email: cdnaldg@nig.nih.gov
              Plate: B0281      row: B      column: 08
              Seq primer: M13 Reverse
              High quality sequence stop: 498
              POLYA=No.
              Location/Qualifiers
                1..498
                  /organism="Mus musculus"
                  /mol_type="mRNA"
                  /strain="C57BL/6J"

```

/db\_xref="nlaEST:B0281B08-5"  
 /db\_xref="taxon:10090"  
 /clone="NIA:B0281B08 IMAGE:30111187"  
 /sex="male"  
 /dev\_stage="embryonic day 8"  
 /lab\_host="DH10B"  
 /clone\_11b="NIA Mouse Embryonic Germ Cell cDNA library (long)"  
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001) [PMID: 11544199]). Total RNAs were obtained from Dr. Mark G. Carter (NIH/NIA-IRP). EG cells were cultured at 37. C, 5% CO2 in DMEM supplemented with 15% ES cell-qualified FBS, 0.1mM non-essential amino acids, 2 mM glutamine, penicillin/streptomycin, and 10000000 units of LIF per liter. beta-mercaptoethanol, and synthesized with an Oligo(dT) primer (Invitrogen): 5'-pGACTAGTCTTAGATCGGAGCGCGCCCTTTT-3' from 2.5 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 4.0 kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 2.1%; Score 22; DB 6; Length 498;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 AGTCCATTTCAGATGAGCA 86  
 ||||||||||||||||  
 Db 462 AGTCCATTTCAGATGAGCA 441

RESULT 7  
 CDS54299 505 bp mRNA linear EST 11-JUN-2003  
 LOCUS B0374A09-5 NIA Mouse E9.5 Whole Embryo cDNA library (long) Mus  
 DEFINITION Musculus cDNA clone NIA:B0374A09 IMAGE:30436136 5', mRNA sequence.  
 ACCESSION CDS54299  
 VERSION CDS54299.1 GI:31602030  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 505)  
 Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.  
 Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method  
 Genome Res. 11 (9), 1553-1558 (2001)

JOURNAL MEDLINE 21429098  
 PUBMED 11544199  
 COMMENT Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Caswell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@lgsun.grc.nia.nih.gov  
 Plate: B0374 row: A column: 09

Seq primer: M13 Reverse  
 High quality sequence stop: 505  
 POLYA=No.  
 FEATURES  
 source  
 Location/Qualifiers  
 1. 505  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="nlaEST:B0374A09-5"  
 /db\_xref="taxon:10090"  
 /clone="NIA:B0374A09 IMAGE:30436136"  
 /tissue\_type="E9.5 whole embryo"  
 /dev\_stage="whole embryo including extraembryonic tissues at 9.5-days postcoitum"  
 /lab\_host="DH10B"  
 /clone\_11b="NIA Mouse E9.5 whole Embryo cDNA library (long)"

/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001) [PMID: 11544199]). Total RNAs were extracted from a pool of 16 embryos at 9.5-days postcoitum. Double-stranded cDNAs were synthesized with an Oligo(dT) primer (Invitrogen): 5'-pGACTAGTCTTAGATCGGAGCGCGCCCTTTT-3' from 1.1 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 2.1%; Score 22; DB 6; Length 505;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 AGTCCATTTCAGATGAGCA 86  
 ||||||||||||||||  
 Db 462 AGTCCATTTCAGATGAGCA 441

RESULT 8  
 CDS28031 744 bp mRNA linear EST 09-JUL-2003  
 LOCUS UI-M-FY0-cfk-p-20-0-UI.r1 NIH BMAP\_FY0 Mus musculus cDNA clone  
 DEFINITION IMAGE:6850917 5', mRNA sequence.  
 ACCESSION CDS28031  
 VERSION CDS28031.1 GI:29361504  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 744)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: gsgaps-remail.nih.gov  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at  
<http://genome.uio.no/distribution/mouse1.html>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

## FEATURES

## source

Seq primer: PYX-5.

## Location/Qualifiers

1. .744  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6850917"  
/issue\_type="whole brain"  
/dev\_stage="embryo 13.5, 14.5, 16.5, 17.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NTH\_BMAP\_FY0"  
/note="Organ: Brain; Vector: PYX-Asc; Site: 1; Ecor I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCAGACAG. This library was created for the University Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

## ORIGIN

Query Match 2.1%; Score 22; DB 6; Length 744;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 AGTCCCATTTTTCAGATGAGA 86  
|||||  
Db 346 AGTCCCATTTTTCAGATGAGA 325

RESULT 9  
BF233030/c 922 bp mRNA linear EST 14-NOV-2000  
LOCUS 602023960F1 NCI\_CGAP\_Li9 Mus musculus cDNA clone IMAGE:415916 5',  
DEFINITION mRNA sequence.  
ACCESSION BF233030  
VERSION BF233030.1 GI:11143008  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 922)  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE NIH-MGC <http://mgs.nci.nih.gov/>.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Jeffrey B. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
[found through the I.M.A.G.E. Consortium/LINL at:](http://image.jnl.gov)  
<http://image.jnl.gov>  
Plate: LAM9436 row: h column: 15  
High quality sequence stop: 622.

## FEATURES

## source

1. .922  
/organism="Mus musculus"  
/mol\_type="mRNA"

## ORIGIN

Query Match 2.1%; Score 22; DB 2; Length 922;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 AGTCCCATTTTTCAGATGAGA 86  
|||||  
Db 560 AGTCCCATTTTTCAGATGAGA 539

RESULT 10  
BU409267 1706 bp mRNA linear EST 29-NOV-2002  
LOCUS 603158026F1 CSEORBL03 Gallus gallus cDNA clone CHEST175g 5', mRNA  
DEFINITION sequence.  
ACCESSION BU409267  
VERSION BU409267.1 GI:25901938  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 1706)  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
TITLE A Comprehensive Collection of Chicken CDNA  
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
MEDLINE 22335534  
PUBMED 12445392  
COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: [Simon.Hubbard@umist.ac.uk](mailto:Simon.Hubbard@umist.ac.uk).

## FEATURES

## source

1. .1706  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="layer and broiler"  
/db\_xref="taxon:9031"  
/clone="CHSR175g8"  
/sex="Male and female"  
/tissue\_type="Chondrocytes isolated from growth plate  
cartilage"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="CSEORBL03"  
/note="Vector: pBluescript II KS(+); Site 1: EcorI;  
Site 2: NotI; Modification of pBluescript II KS(+)  
[Stratagene] vector to accommodate cDNA produced with the  
T-timed protocol (Construction of uni-directionally  
cloned cDNA libraries from messenger RNA for improved 3'  
end DNA sequencing by Glenn Fu, et al. U.S. Patent #  
5,387,624). Cut pBluescript II KS(+) with NotI and EcorI.  
Ligate in double stranded adaptor containing BspI and  
BamHI sites [5'ggccgcgtgcagccgcgagtcgaaaaag]  
[5'aattcttttcgagtcgcggcgtcgcagc]"

## ORIGIN

Query Match 2.1%; Score 22; DB 5; Length 1706;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 TGGTGTGTCAGCCCGACGCC 182  
DB 255 TGGTGTGTCAGCCCGACGCC 276

RESULT 11  
AK079109 1993 bp mRNA linear HTC 03-APR-2004  
LOCUS AK079109.1  
DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region and neck cDNA, RIKEN full-length enriched library, clone:9430008C03  
ACCESSION AK079109  
VERSION AK079109.1 GI:26098280  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636

REFERENCE  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159

REFERENCE  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Teshito, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861

REFERENCE  
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
MEDLINE 11076861

REFERENCE  
AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
MEDLINE 11076861

REFERENCE  
AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Asekawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Katsunai, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sasaki, C., Sasaki, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, Y., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

FEATURES  
source  
1..1993  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM:DB:9430008C03"  
/db\_xref="taxon:10090"  
/clone="9430008C03"  
/cissue\_type="embryonic body between diaphragm region and neck"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
1..1993  
/note="unknown EST (GB|BI251132, evidence: BLASTN, 98%, match=822)"

misc\_feature  
1..1993  
/note="unknown EST (GB|BI251132, evidence: BLASTN, 98%, match=822)"

ORIGIN

Query Match 2.1%; Score 22; DB 3; Length 1993;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 AGTCCATTTCAGATGAGA 86  
DB 634 AGTCCATTTCAGATGAGA 613

RESULT 12  
AK040893 3268 bp mRNA linear HTC 03-APR-2004  
LOCUS AK040893.1  
DEFINITION Mus musculus adult male aorta and vein cDNA, RIKEN full-length enriched library, clone:A530040I21 product:unknown EST, full insert sequence.  
ACCESSION AK040893  
VERSION AK040893.1 GI:26088138  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636

REFERENCE  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159

REFERENCE  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Teshito, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,



Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubaki, M., Yoneda, Y., Iehikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN Integrated sequence analysis (RISA) system--384-format  
Sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL MEDLINE 20530913  
PUBMED 11076861

4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

JOURNAL REFERENCE  
AUTHORS

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

JOURNAL REFERENCE  
AUTHORS

6 (bases 1 to 3268)  
Adachi, J., Aizawa, K., Akimura, T., Arawaka, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imocani, K., Iehi, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohtsuo, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shitaki, T., Sugabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
Location/Qualifiers

FEATURES  
SOURCE

1..3268  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:A530040121"  
/db\_xref="taxon:10090"  
/clone="A530040121"  
/sex="male"  
/tissue\_type="aorta and vein"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
1..3268  
/note="unknown EST (GB|BI251132, evidence: BLASTN, 98%, match=822)"

ORIGIN

Query Match 2.1%; Score 22; DB 3; Length 3268;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 AGTCCATTTCAGATGAGA 86  
|||||  
625 AGTCCATTTCAGATGAGA 604

RESULT 13  
BI25065/c 204 bp mRNA linear EST 30-JUL-2001  
LOCUS 602980483F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:5133309 5',  
DEFINITION mRNA sequence.  
ACCESSION BI25065  
VERSION BI25065.1 GI:15013722  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.  
1 (bases 1 to 204)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: gcgaps@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at:  
http://image.llnl.gov  
Plate: LMNL1327 row: e column: 22  
High quality sequence stop: 204.  
Location/Qualifiers

FEATURES  
SOURCE

1..204  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5133309"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP\_L19"  
/note="Organ: liver; Vector: pCMV-SPORT6, Site 1: NciI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

ORIGIN

Query Match 2.0%; Score 21; DB 4; Length 204;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 903 TGTGGGCTGAGGCTGGCTG 923  
|||||  
48 TGTGGGCTGAGGCTGGCTG 28

RESULT 14  
CR516621 207 bp mRNA linear EST 07-JUL-2004  
LOCUS CR516621 Mouse pBluescript Lion Mus musculus cDNA clone  
DEFINITION LI0NP462B0815 3', mRNA sequence.  
ACCESSION CR516621  
VERSION CR516621.1 GI:49908681  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.  
1 (bases 1 to 207)  
Henrich, J., Hermans, J., Kranz, H., Loebbert, R., Schlueter, T., Schuette, D., Weindel, M., Hell, O., Ebert, L., Newbert, P., Peters, M., Radehof, U., Schneider, D. and Korn, B.  
Mouse ArrayTNG cDNA  
Unpublished (2004)  
Contact: Inge Airlart  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2005, 18:13:24 ; Search time 219 Seconds  
(without alignments)  
7755.503 Million cell updates/sec

Title: US-10-617-443B-1\_COPY\_1\_1038

Perfect score: 1038

Sequence: 1 caggagactcgaagcttcgag.....ggcccttcgcccgagcgag 1038

Scoring table: OLIGO NUC

Gapop 60.0, Gapext 60.0

Searched: 1202784 seqs, 818138359 residues

Word size: 12

Total number of hits satisfying chosen parameters: 83567

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

- 1: Issued Patents NA:\*
- 2: /cgn2\_6/ptodaca/1/ina/5A COMB.seq:\*
- 3: /cgn2\_6/ptodaca/1/ina/5B COMB.seq:\*
- 4: /cgn2\_6/ptodaca/1/ina/6A COMB.seq:\*
- 5: /cgn2\_6/ptodaca/1/ina/6B COMB.seq:\*
- 6: /cgn2\_6/ptodaca/1/ina/PTCUS COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	2.1	116652	4 US-09-949-016-13413	Sequence 13413, A
2	21	2.0	65744	4 US-09-949-016-12591	Sequence 12591, A
3	21	2.0	65745	4 US-09-949-016-15871	Sequence 15871, A
4	21	2.0	818128	4 US-09-949-016-14546	Sequence 14546, A
5	21	2.0	818128	4 US-09-949-016-14547	Sequence 14547, A
6	21	2.0	818128	4 US-09-949-016-14548	Sequence 14548, A
7	21	2.0	818128	4 US-09-949-016-14549	Sequence 14549, A
8	21	2.0	818128	4 US-09-949-016-14550	Sequence 14550, A
9	21	2.0	818128	4 US-09-949-016-14551	Sequence 14551, A
10	21	2.0	818128	4 US-09-949-016-14552	Sequence 14552, A
11	21	2.0	818128	4 US-09-949-016-14553	Sequence 14553, A
12	21	2.0	818128	4 US-09-949-016-14554	Sequence 14554, A
13	21	2.0	818128	4 US-09-949-016-14555	Sequence 14555, A
14	21	2.0	818128	4 US-09-949-016-14556	Sequence 14556, A
15	21	2.0	818128	4 US-09-949-016-14557	Sequence 14557, A
16	21	2.0	818128	4 US-09-949-016-14558	Sequence 14558, A
17	21	2.0	818128	4 US-09-949-016-14559	Sequence 14559, A
18	21	2.0	818128	4 US-09-949-016-14560	Sequence 14560, A
19	21	2.0	818128	4 US-09-949-016-14561	Sequence 14561, A
20	21	2.0	818128	4 US-09-949-016-14562	Sequence 14562, A
21	21	2.0	818128	4 US-09-949-016-14563	Sequence 14563, A
22	21	2.0	818128	4 US-09-949-016-14564	Sequence 14564, A
23	21	2.0	818128	4 US-09-949-016-14565	Sequence 14565, A
24	21	2.0	818128	4 US-09-949-016-14566	Sequence 14566, A
25	20	1.9	601	4 US-09-949-016-14567	Sequence 14567, A
26	20	1.9	601	4 US-09-949-016-14568	Sequence 14568, A
27	20	1.9	17353	4 US-09-949-016-13437	Sequence 13437, A

28	20	1.9	28555	4 US-09-949-016-13106	Sequence 13106, A
29	19	1.8	89	4 US-09-513-999C-16411	Sequence 16411, A
30	19	1.8	500	4 US-09-866-108A-15742	Sequence 15742, A
31	19	1.8	558	1 US-08-244-492A-4	Sequence 4, Appl
32	19	1.8	558	1 US-08-709-915-5	Sequence 5, Appl
33	19	1.8	601	4 US-09-949-016-26845	Sequence 26845, A
34	19	1.8	601	4 US-09-949-016-47781	Sequence 47781, A
35	19	1.8	601	4 US-09-949-016-55840	Sequence 55840, A
36	19	1.8	601	4 US-09-949-016-56305	Sequence 56305, A
37	19	1.8	601	4 US-09-949-016-155557	Sequence 155557, A
38	19	1.8	601	4 US-09-949-016-155656	Sequence 155656, A
39	19	1.8	637	4 US-09-517-431E-18	Sequence 18, Appl
40	19	1.8	1215	3 US-09-091-952A-29	Sequence 29, Appl
C 41	19	1.8	25202	4 US-09-949-016-13151	Sequence 13151, A
C 42	19	1.8	27380	4 US-09-949-016-12877	Sequence 12877, A
C 43	19	1.8	27383	4 US-09-949-016-14393	Sequence 14393, A
C 44	19	1.8	38078	4 US-09-949-016-12429	Sequence 12429, A
45	19	1.8	38084	4 US-09-949-016-16459	Sequence 16459, A

# ALIGNMENTS

```

RESULT 1
; Sequence 13413, Application US/09949016
; Patent No. 6812335
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13413
; LENGTH: 116652
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(116652)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13413
Query Match 2.1%; Score 22; DB 4; Length 116652;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
Cy 931 ACAGGGGCTCAGGAAGGCT 952
Db 75459 ACAGGGGCTCAGGAAGGCT 75438
;
RESULT 2
US-09-949-016-12591
; Sequence 12591, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

```

```

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12591
; LENGTH: 65744
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(65744)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12591
```

```

Query Match      2.0%; Score 21; DB 4; Length 65744;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      124 CACACAGCTAGAAAGCAGCCA 144
Db      44923 CACACAGCTAGAAAGCAGCCA 44943
```

```

RESULT 3
US-09-949-016-15871
; Sequence 15871, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15871
; LENGTH: 65745
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(65745)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15871
```

```

Query Match      2.0%; Score 21; DB 4; Length 65745;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      124 CACACAGCTAGAAAGCAGCCA 144
Db      44923 CACACAGCTAGAAAGCAGCCA 44943
```

```

RESULT 4
US-09-949-016-14546
; Sequence 14546, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
```

```

; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14546
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14546
```

```

Query Match      2.0%; Score 21; DB 4; Length 818128;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      120 AGGCCACAGCTAGAAAGCA 140
Db      116875 AGGCCACAGCTAGAAAGCA 116895
```

```

RESULT 5
US-09-949-016-14547
; Sequence 14547, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14547
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14547
```

```

Query Match      2.0%; Score 21; DB 4; Length 818128;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      120 AGGCCACAGCTAGAAAGCA 140
Db      116875 AGGCCACAGCTAGAAAGCA 116895
```

```

RESULT 6
US-09-949-016-14548
; Sequence 14548, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14548
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14548

```

```

Query Match
Best Local Similarity 100.0%; Score 21; DB 4; Length 818128;
Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 120 AGGCCACACAGCTAGAAAGCA 140
Db 116875 AGGCCACACAGCTAGAAAGCA 116895

```

```

RESULT 7
US-09-949-016-14549
; Sequence 14549, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14549
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14549

```

```

Query Match
Best Local Similarity 100.0%; Score 21; DB 4; Length 818128;
Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 120 AGGCCACACAGCTAGAAAGCA 140
Db 116875 AGGCCACACAGCTAGAAAGCA 116895

```

```

RESULT 8
US-09-949-016-14550
; Sequence 14550, Application US/09949016

```

```

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14550
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14550

```

```

Query Match
Best Local Similarity 100.0%; Score 21; DB 4; Length 818128;
Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 120 AGGCCACACAGCTAGAAAGCA 140
Db 116875 AGGCCACACAGCTAGAAAGCA 116895

```

```

RESULT 9
US-09-949-016-14551
; Sequence 14551, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14551
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14551

```

```

Query Match
Best Local Similarity 100.0%; Score 21; DB 4; Length 818128;
Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 120 AGGCCACACAGCTAGAAAGCA 140
Db 116875 AGGCCACACAGCTAGAAAGCA 116895

```

```
RESULT 10
US-09-949-016-14552
/ Sequence 14552, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 14552
/ LENGTH: 818128
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(818128)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14552
```

```
Query Match          2.0%; Score 21; DB 4; Length 818128;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      120 AGGCCACACAGCTAGAAAGCA 140
      |||||||
Db      116875 AGGCCACACAGCTAGAAAGCA 116895
```

```
RESULT 11
US-09-949-016-14553
/ Sequence 14553, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 14553
/ LENGTH: 818128
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(818128)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14553
```

```
Query Match          2.0%; Score 21; DB 4; Length 818128;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      120 AGGCCACACAGCTAGAAAGCA 140
      |||||||
```

```
Db      116875 AGGCCACACAGCTAGAAAGCA 116895
```

```
RESULT 12
US-09-949-016-14554
/ Sequence 14554, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 14554
/ LENGTH: 818128
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(818128)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14554
```

```
Query Match          2.0%; Score 21; DB 4; Length 818128;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      120 AGGCCACACAGCTAGAAAGCA 140
      |||||||
Db      116875 AGGCCACACAGCTAGAAAGCA 116895
```

```
RESULT 13
US-09-949-016-14555
/ Sequence 14555, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 14555
/ LENGTH: 818128
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(818128)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14555
```

```
Query Match          2.0%; Score 21; DB 4; Length 818128;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 120 AGGCCACACAGCTAGAAAGCA 140  
 |||||  
 Db 116875 AGGCCACACAGCTAGAAAGCA 116895

RESULT 14

US-09-949-016-14556  
 ; Sequence 14556, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CLO01307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 14556  
 ; LENGTH: 818128  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(818128)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-14556

Query Match 2.0%; Score 21; DB 4; Length 818128;  
 Best Local Similarity 100.0%; Pred. No. 2.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 120 AGGCCACACAGCTAGAAAGCA 140  
 |||||  
 Db 116875 AGGCCACACAGCTAGAAAGCA 116895

RESULT 15

US-09-949-016-14557  
 ; Sequence 14557, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CLO01307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 14557  
 ; LENGTH: 818128  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(818128)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-14557

Query Match 2.0%; Score 21; DB 4; Length 818128;  
 Best Local Similarity 100.0%; Pred. No. 2.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 120 AGGCCACACAGCTAGAAAGCA 140  
 |||||  
 Db 116875 AGGCCACACAGCTAGAAAGCA 116895

Search completed: June 29, 2005, 21:11:05  
 Job time : 234 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6.  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2005, 15:40:04 ; Search time 675 Seconds  
(without alignments)  
9103.247 Million cell updates/sec

Title: US-10-617-443b-1\_COPY\_1\_1038  
Perfect score: 1038  
Sequence: 1 cagggaactcgaagcttcgag.....ggccctcccgccagcgcag 1038

Scoring table: ORIGO NUC  
Gapop 60.0, Gapext 60.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 12

Total number of hits satisfying chosen parameters: 436444

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1038	100.0	3006	12 ADJ11355	Adj11355 Human PRS
2	22	2.1	1379	3 AAF21695	Aaf21695 Human bre
3	22	2.1	4558	10 ADD47518	Add47518 Human gen
4	22	2.1	4558	10 ADD47514	Add47514 Human gen
5	22	2.1	112453	13 ABD32665	Abd32665 Human can
6	22	2.0	34	12 ADJ11359	Adj11359 PCR prime
7	21	2.0	244	12 ADG39925	Adg39925 Kidney dt
8	21	2.0	1795	6 ABL98890	Ab198890 Human sec
9	21	2.0	3157	12 ADG87121	Adg87121 Human tum
10	21	2.0	3983	8 ABZ36074	Abz36074 Human sec
11	21	2.0	4030	12 ADM87103	Adm87103 Human pro
12	21	2.0	127145	13 ADQ80254	Adq80254 Hermansky
13	21	2.0	189013	8 ACF62741	Acf62741 Cancer ba
14	21	2.0	189013	8 ADB20856	Adb20856 MRP1 base
15	21	2.0	189013	10 ADB87945	Adb87945 Human UGT
16	21	2.0	189013	10 ADB86928	Adb86928 Human MDR
17	21	2.0	189013	10 ADB82119	Adb82119 Human MDR
18	20	1.9	20	12 ADJ11361	Adj11361 PCR prime
19	20	1.9	301	4 AAH57233	Aah57233 Human int
20	20	1.9	836	2 AAQ06484	Aaq06484 Intestina

## ALIGNMENTS

RESULT 1	ID	ADJ11355	standard; cDNA; 3006 BP.
XX	ADJ11355;		
DT	20-MAY-2004	(first entry)	
XX			
DE	Human PRS11-L cDNA encoding a serine protease Segid 1.		
XX			
KW	human; PRS11-L; Gene; ss; S2 serine protease; S2/HR74;		
KW	cellular physiology; apoptosis; gene therapy; osteoarthritis; stress;		
KW	apoptotic; osteopathic; antitachytic; tranquiliser.		
XX			
OS	Homo sapiens.		
XX			
FT	Key.	Location/Qualifiers	
FT	CDS	1011..2015	
FT		/*tag= a	
FT		/product= "PRS11-L protein"	
XX	US2004005659-A1.		
XX	08-JAN-2004.		
XX			
PF	03-JUL-2002; 2002US-00189099.		
XX			
PR	03-JUL-2002; 2002US-00189099.		
XX			
PA	(DARR/) DARROW A L.		
PA	(OLJT/) OL J.		
PA	(CHEN/) CHEN C.		
PA	(ANDR/) ANDRADE-GORDON P.		
PI			
XX	Darrow AL, OL J, Chen C, Andrade-Gordon P;		
DR	WPI; 2004-081723/08.		
XX	P-PSDB; ADJ11356.		
XX			
PT	New isolated S2 serine protease nucleic acids and polypeptides, useful		
PT	for diagnosing and/or treating diseases with aberrant expression or		
PT	activity the S2 serine protease, such as osteoarthritis, stress and		
PT	apoptotic disorders.		

C 21	20	1.9	836	2	AAQ12218	Aaq12218 SMUC-41 i
C 22	20	1.9	3294	4	AAH57441	Aah57441 Human int
C 23	20	1.9	15720	4	AAI29504	Aai29504 C899P det
C 24	20	1.9	15720	6	ABL62358	Ab162358 Colon ade
C 25	20	1.9	15720	8	ABZ33690	Abz33690 Human col
C 26	20	1.9	15720	8	ACD13431	Acd13431 Human DNA
C 27	20	1.9	15720	10	AAD63141	Aad63141 Human muc
C 28	20	1.9	15720	10	AAD62742	Aad62742 Human muc
C 29	20	1.9	15720	11	ADP88237	Adp88237 Colon can
C 30	20	1.9	15720	11	ADP88236	Adp88236 Colon can
C 31	20	1.9	15720	12	ADQ29628	Adq29628 Human col
C 32	20	1.9	15720	13	ADQ80264	Adq80264 Intestina
C 33	20	1.9	26000	13	ADR86701	Adr86701 Human Eph
C 34	20	1.9	26000	13	ADR86701	Adr86701 Human Eph
C 35	20	1.9	55875	12	ADQ59503	Adq59503 Human can
C 36	20	1.9	79860	11	ACN44122	Actn44122 Human gen
C 37	20	1.9	87761	11	ADL27164	Adl27164 Human gen
C 38	20	1.9	94529	9	ADA03086	Ada03086 Human hCG
C 39	20	1.9	94529	10	ADB72824	Adb72824 Human hCG
C 40	20	1.9	94531	9	ADA66370	Ada66370 Human hCG
C 41	20	1.9	117962	8	AAD54480	Aad54480 Human CIP
C 42	19	1.8	89	3	AAC12336	Act12336 Human sec
C 43	19	1.8	251	12	ADM72179	Adm72179 Human NTR
C 44	19	1.8	254	7	ADS65486	Ads65486 Corn seed
C 45	19	1.8	290	6	ABV95882	Abv95882 Human pan

XX Claim 2; SEQ ID NO 1; 28pp; English.  
 PS This invention relates to a novel isolated nucleic acid molecule  
 CC identified as PRS11-L that encodes an S2 serine protease. Specifically,  
 CC it refers to members of the S2/Htra serine protease family, such that it  
 CC plays a role in cellular physiology and apoptosis. The present invention  
 CC provides agonists, antagonists, antibodies and recombinant expression  
 CC vectors useful in methods of treatment, or detection and diagnosis of  
 CC diseases associated with the aberrant expression or activity of the S2  
 CC serine protease, PRS11-L. Accordingly, compositions described herein can  
 CC be used via gene therapy routes to treat osteoarthritis, stress and  
 CC apoptotic disorders. As such, they exhibit osteoprotective, antiarthritic and  
 CC transglutinin activities. This polynucleotide sequence is the human PRS11  
 CC -L cDNA sequence of the invention.  
 XX  
 SQ Sequence 3006 BP; 615 A; 918 C; 911 G; 562 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 1038; DB 12; Length 3006;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1038; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGGGAATCGAATGTTGACAGTCTCCACACTGATCCACAGATGTGTAGAGGGCAT 60  
 DB 1 CAGGGAATCGAATGTTGACAGTCTCCACACTGATCCACAGATGTGTAGAGGGCAT 60  
 QY ATTCAATCCCATTTTTCAGATGAGAGTTGAGGCCCAAGAGAACTAATATCTGTCTGA 120  
 DB ATTCAATCCCATTTTTCAGATGAGAGTTGAGGCCCAAGAGAACTAATATCTGTCTGA 120  
 QY 61 ATTCAATCCCATTTTTCAGATGAGAGTTGAGGCCCAAGAGAACTAATATCTGTCTGA 120  
 DB 61 ATTCAATCCCATTTTTCAGATGAGAGTTGAGGCCCAAGAGAACTAATATCTGTCTGA 120  
 QY 121 GGCCACACAGCTGAAAGCAGCCAGCCAGCCAAACCCCTGTGTGTGACAGCCCCACAG 180  
 DB 121 GGCCACACAGCTGAAAGCAGCCAGCCAGCCAAACCCCTGTGTGTGACAGCCCCACAG 180  
 QY 181 CCAATGCTCTATTCGCGGCTCGGAGCCACGAGCAGAGCTGACAGCATGTGTCTCA 240  
 DB 181 CCAATGCTCTATTCGCGGCTCGGAGCCACGAGCAGAGCTGACAGCATGTGTCTCA 240  
 QY 241 TGGTGGGAATCGAGAGAGCCCGGACAGCCCGTGTGAGGAAACCCGAGGGCTGTAGGC 300  
 DB 241 TGGTGGGAATCGAGAGAGCCCGGACAGCCCGTGTGAGGAAACCCGAGGGCTGTAGGC 300  
 QY 301 CCCGTGCACTGATGCTCTGAGGCTGTGTGCTCTGAGCAGCCAGCCCTACTGTGACG 360  
 DB 301 CCCGTGCACTGATGCTCTGAGGCTGTGTGCTCTGAGCAGCCAGCCCTACTGTGACG 360  
 QY 361 GCAGCAGAAATCTGAGCCCGGAGGAGGATCCAGGAAAGTTCTGTAAACATCTAGCAAGTCG 420  
 DB 361 GCAGCAGAAATCTGAGCCCGGAGGAGGATCCAGGAAAGTTCTGTAAACATCTAGCAAGTCG 420  
 QY 421 GGCTGGGGTGTGGCCAAATGACACAGATGTAGGGCCCTGTGAGCTCAAGAAATTTGGCAG 480  
 DB 421 GGCTGGGGTGTGGCCAAATGACACAGATGTAGGGCCCTGTGAGCTCAAGAAATTTGGCAG 480  
 QY 481 CTCTTTTGGCCAGAGGGGCGCAGCTGTGTCCGGGCTGTGGGTAGCTCAGAAAGGTCACCT 540  
 DB 481 CTCTTTTGGCCAGAGGGGCGCAGCTGTGTCCGGGCTGTGGGTAGCTCAGAAAGGTCACCT 540  
 QY 541 GGGGGTCTTTCATACACCCCGCTGTGACACTGTGTAGCCCGGAGGCTCGAGAGGACCC 600  
 DB 541 GGGGGTCTTTCATACACCCCGCTGTGACACTGTGTAGCCCGGAGGCTCGAGAGGACCC 600  
 QY 601 AGCTGAGGCCAATGAGAGAGGGCCAGTCTCTCTGTAAAGGATTTGCTGTAGCAGTAG 660  
 DB 601 AGCTGAGGCCAATGAGAGAGGGCCAGTCTCTCTGTAAAGGATTTGCTGTAGCAGTAG 660  
 QY 661 GGAACAGACAGGCCCGAGGGGAGCTAAACCGAGATCCAGCCCGGCTCACTCCCTGTG 720  
 DB 661 GGAACAGACAGGCCCGAGGGGAGCTAAACCGAGATCCAGCCCGGCTCACTCCCTGTG 720  
 QY 721 GCTCAGGCAATATCTTAACCTCTCTGTAGGCTCTCTGAGCCAGCTTACAGAGGATTCAGTG 780  
 DB 721 GCTCAGGCAATATCTTAACCTCTCTGTAGGCTCTCTGAGCCAGCTTACAGAGGATTCAGTG 780

QY 781 AGGGGGGTGAGAGAGCCAGCAGCGTGAAGCCTTTTAACCATCTCGGGGTGAGCAGC 840  
 DB 781 AGGGGGGTGAGAGAGCCAGCAGCGTGAAGCCTTTTAACCATCTCGGGGTGAGCAGC 840  
 QY 841 CCTTCCCAATGCTGTGTCTGACTGACACTGCTGTGTGTGTAGAGGGGTCCCAACGGGCTC 900  
 DB 841 CCTTCCCAATGCTGTGTCTGACTGACACTGCTGTGTGTGTAGAGGGGTCCCAACGGGCTC 900  
 QY 901 AGTGTGGGTGTAGGCTGTGTCTGAACTGGAGAGAGGGGTCTCAGAGAAAGCCCTCTCTCC 960  
 DB 901 AGTGTGGGTGTAGGCTGTGTCTGAACTGGAGAGAGGGGTCTCAGAGAAAGCCCTCTCTCC 960  
 QY 961 TCCCACTGGGAGTATGAGCCCTGTGGAGCTGGAGAGCATGTGATCTCACTGATGACACTGG 1020  
 DB 961 TCCCACTGGGAGTATGAGCCCTGTGGAGCTGGAGAGCATGTGATCTCACTGATGACACTGG 1020  
 QY 1021 CCTTCCCGCCAGCCGAG 1038  
 DB 1021 CCTTCCCGCCAGCCGAG 1038

RESULT 2  
 AAF21695/c  
 ID AAF21695 standard; DNA; 1379 BP.  
 XX  
 AC AAF21695;  
 XX  
 DT 27-MAR-2001 (first entry)  
 XX  
 DE Human breast and ovarian cancer associated antigen gene SEQ ID 82.  
 XX  
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
 KW neotropic; neuroprotective; antiviral; antiatheric; hepatotropic;  
 KW antiatheric; antiinflammatory; antitumor; vulnary; anticonvulsant;  
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;  
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
 KW cardiovascular disorder; wound healing; neurological disease; de.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO20055173-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000MO-US005881.  
 XX  
 PR 12-MAR-1999; 99US-0124270P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2000-611515/58.  
 DR P-PSDB; AAB58792.  
 XX  
 PT New human breast and ovarian cancer associated gene sequences and the  
 PT polypeptides encoded by these genes, useful in the prevention, treatment  
 PT and diagnosis of cancer, immune disorders, cardiovascular disorders and  
 PT neurological diseases.  
 XX  
 PS Claim 1, Page 543; 1299pp; English.  
 XX  
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
 CC associated with breast and ovarian cancer. Included in the invention are  
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
 CC isolation and characterisation of the DNA and protein sequences of the  
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
 CC or antagonist sequences exhibit cytostatic; immunosuppressive; neotropic;  
 CC neuroprotective; antiviral; antiatheric; hepatotropic; antiatheric;

CC antiinflammatory; antitumor; anticonvulsant; antibacterial;  
 CC antifungal; antiparasitic and cardiant activity. The polynucleotide and  
 CC protein sequences are used in the diagnosis of cancer, particularly  
 CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists  
 CC and agonists may also be used in the diagnosis, prevention and treatment  
 CC of immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
 CC cardiovascular disorders such as myocardial ischaemia; wound healing;  
 CC neurological diseases such as cerebral anoxia and epilepsy; and  
 CC infectious diseases  
 CC  
 SQ Sequence 1379 BP; 301 A; 354 C; 385 G; 330 T; 0 U; 9 Other;  
 XX  
 Query Match 2.1%; Score 22; DB 3; Length 1379;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 120 AGGCACACAGCTGAGAAAGCAG 141  
 Db 790 AGGCACACAGCTGAGAAAGCAG 769  
 XX  
 RESULT 3  
 ADD47518  
 ID ADD47518 standard; DNA; 4558 BP.  
 XX  
 AC ADD47518;  
 XX  
 DT 02-DEC-2004 (revised)  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human gene XM\_045474, SEQ ID NO 13214.  
 XX  
 KM Human; de; gene; pain; neuronal tissue; gene therapy;  
 KM spinal segmental nerve injury; chronic constriction injury; CCI;  
 KM spared nerve injury; SNI; Chung.  
 XX  
 OS Homo sapiens.  
 OS Unidentified.  
 XX  
 PN WO2003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 PI Woolf C, D'Urso D, Befort K, Costigan M;  
 XX  
 DR WPI; 2003-268312/26.  
 DR GENBANK; XM\_045474.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 PS Example 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal

CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human DNA (described in Table 3 of  
 CC the specification) which encodes one of the polypeptides of the invention  
 CC which is differentially expressed during pain. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 4558 BP; 1030 A; 1062 C; 1303 G; 1163 T; 0 U; 0 Other;  
 XX  
 Query Match 2.1%; Score 22; DB 10; Length 4558;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 772 GGTCCACTGAGGGGGGTGAGGA 793  
 Db 4145 GGTCCACTGAGGGGGGTGAGGA 4166  
 XX  
 RESULT 4  
 ADD47514  
 ID ADD47514 standard; DNA; 4558 BP.  
 XX  
 AC ADD47514;  
 XX  
 DT 02-DEC-2004 (revised)  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human gene XM\_045474, SEQ ID NO 13210.  
 XX  
 KM Human; de; gene; pain; neuronal tissue; gene therapy;  
 KM spinal segmental nerve injury; chronic constriction injury; CCI;  
 KM spared nerve injury; SNI; Chung.  
 XX  
 OS Homo sapiens.  
 OS Unidentified.  
 XX  
 PN WO2003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 PI Woolf C, D'Urso D, Befort K, Costigan M;  
 XX  
 DR WPI; 2003-268312/26.  
 DR GENBANK; XM\_045474.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 PS Example 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat

or human polynucleotides or a polynucleotide which represents a fragment, CC  
CC derivative or allelic variation of the nucleic acid sequence. Also CC  
CC claimed are a vector comprising the novel polynucleotide, a host cell CC  
CC comprising the vector, a method for identifying a nucleotide sequence CC  
CC which is differentially regulated in an animal subjected to pain and a CC  
CC kit to perform the method, an array, a method for identifying an agent CC  
CC that increases or decreases the expression of the polynucleotide sequence CC  
CC that is differentially expressed in neuronal tissue of a first animal CC  
CC subjected to pain, a method for identifying a compound which regulates CC  
CC the expression of a polynucleotide sequence which is differentially CC  
CC expressed in an animal subjected to pain, a method for identifying a CC  
CC compound that regulates the activity of one or more of the CC  
CC polynucleotides, a method for producing a pharmaceutical composition, a CC  
CC method for identifying a compound or small molecule that regulates the CC  
CC activity in an animal of one or more of the polypeptides given in the CC  
CC specification, a method for identifying a compound useful in treating CC  
CC pain and a pharmaceutical composition comprising the one or more CC  
CC polypeptides or their antibodies. The polynucleotide or the compound that CC  
CC modulates its activity is useful for preparing a medicament for treating CC  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction CC  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene CC  
CC therapy). The sequence presented is a human DNA (described in Table 3 of CC  
CC the specification) which encodes one of the polypeptides of the invention CC  
CC for this patent did not form part of the printed specification, but was CC  
CC obtained in electronic form directly from WIPO at CC  
CC ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 4558 BP; 1030 A; 1062 C; 1303 G; 1163 T; 0 U; 0 Other;

Query Match 2.1%; Score 22; DB 10; Length 4558;

Best Local Similarity 100.0%; Pred. No. 5.2; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 772 GGTCCACTGAGCGGGGTGAGCA 793

Db 4145 GGTCCACTGAGCGGGGTGAGCA 4166

RESULT 5

ABD32665/c

ID ABD32665 standard; DNA; 112453 BP.

XX ABD32665;

DT 18-NOV-2004 (first entry)

DE Human cancer-associated genomic DNA HD13-070.

XX Human; ds; cancer-associated protein; gene; cytostatic; cancer;

KW Leukaemia; lymphoma; CAP.

XX Homo sapiens.

OS WO2004074320-A2.

PN 02-SEP-2004.

PD 17-FEB-2004; 2004WO-US004730.

PF 14-FEB-2003; 2003US-00367094.

PR 14-MAR-2003; 2003US-00388838.

PR 15-APR-2003; 2003US-00417375.

PR 13-JUN-2003; 2003US-00461862.

PR 15-SEP-2003; 2003US-00663431.

PR 15-DEC-2003; 2003US-00737318.

XX (SAGR-) SAGRES DISCOVERY INC.

XX Morris DW, Morris DW, Malandro MS,

XX MPI, 2004-652914/63.

PT New isolated cancer-associated polynucleotides and polypeptides useful CC  
PT for diagnosing, preventing or treating cancers, especially lymphoma and CC  
PT leukemia, or in screening for agents that modulate cancer. CC  
XX claim 16; seqid 209; 310pp; English.

PS The invention relates to an isolated nucleic acid comprising at least 10 CC  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given CC  
CC in the specification, or its complement. The nucleic acids encode cancer- CC  
CC associated proteins. Also included are an expression vector comprising CC  
CC the isolated nucleic acid cited above, a host cell comprising the above CC  
CC recombinant nucleic acid or expression vector, a microarray for detecting CC  
CC a cancer-associated (CA) nucleic acid comprising at least one probe CC  
CC comprising at least 10 contiguous nucleotides of any of the above- CC  
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within CC  
CC an open reading frame of a CA sequence selected from any of the 95 CC  
CC polynucleotide sequences as mentioned in the specification, or its CC  
CC complement), an isolated antibody, (or its antigen binding fragment) that CC  
CC binds to the above polypeptide, a hybridoma that produces the above CC  
CC monoclonal antibody, a pharmaceutical composition comprising the above CC  
CC antibody and a pharmaceutical excipient, a kit for detecting cancer CC  
CC cells (comprising the antibody cited above, methods for diagnosing cancer CC  
CC or for detecting the presence or absence of cancer cells in an CC  
CC individual, a method for inhibiting growth of cancer cells in an CC  
CC individual, a method for delivering a therapeutic agent to cancer cells CC  
CC in an individual, an electronic library comprising the above CC  
CC polynucleotide or polypeptide (or their fragments), methods of screening CC  
CC for anticancer activity or for a bioactive agent capable of modulating CC  
CC the activity of a CA protein (CAP), methods for detecting cancer CC  
CC associated with expression of a polypeptide in a test cell sample, a CC  
CC method for treating cancers and a method for inhibiting the expression of CC  
CC CA gene in a cell. The composition and methods are useful for detecting, CC  
CC diagnosing, preventing and treating cancers, especially lymphoma and CC  
CC leukaemia. These may also be used in screening for agents that modulate CC  
CC cancer. The present sequence is a human CAP genomic sequence. Note: The CC  
CC sequence data for this patent did not form part of the printed CC  
CC specification, but was obtained in electronic format directly from WIPO CC  
CC at ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 112453 BP; 28299 A; 24810 C; 27274 G; 32070 T; 0 U; 0 Other;

Query Match 2.1%; Score 22; DB 13; Length 112453;

Best Local Similarity 100.0%; Pred. No. 4.6; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 CTGCTGAGCCACACAGCTAG 134

Db 33503 CTGCTGAGCCACACAGCTAG 33482

RESULT 6

ADJ11359

ID ADJ11359 standard; DNA; 34 BP.

XX ADJ11359;

DT 20-MAY-2004 (first entry)

DE PCR primer used to construct the human PRSS11-L expression vector SeqID5.

XX human; PRSS11-L; PCR; ss; S2 serine protease; S2/HTRA;

KW cellular physiology; apoptosis; gene therapy; osteoarthritis; stress;

KW apoptotic; osteopathic; antiarthritic; tranquiliser; primer.

XX Homo sapiens.

OS Synthetic.

PN US2004005659-A1.

XX 08-JAN-2004.

XX 03-JUL-2002; 2002US-00189099.

PR 03-JUN-2002; 2002US-00189099.  
 XX (DARR/) DARROW A L.  
 PA (OLJ/) OL J.  
 PA (CHEN/) CHEN C.  
 PA (ANDR/) ANDRADE-GORDON P.  
 PI Darrow AL, OI J, Chen C, Andrade-Gordon P;  
 XX WPI; 2004-081723/08.  
 XX  
 XX New isolated S2 serine protease nucleic acids and polypeptides, useful  
 PT for diagnosing and/or treating diseases with aberrant expression or  
 PT activity the S2 serine protease, such as osteoarthritis, stress and  
 PT apoptotic disorders.  
 XX  
 XX Example 3; SEQ ID NO 5; 28pp; English.  
 XX  
 CC This invention relates to a novel isolated nucleic acid molecule  
 CC identified as PRS51-L that encodes an S2 serine protease. Specifically,  
 CC it refers to members of the S2/HTRA serine protease family, such that it  
 CC plays a role in cellular physiology and apoptosis. The present invention  
 CC provides agonists, antagonists, antibodies and recombinant expression  
 CC vectors useful in methods of treatment, or detection and diagnosis of  
 CC diseases associated with the aberrant expression or activity of the S2  
 CC serine protease, PRS51-L. Accordingly, compositions described herein can  
 CC be used via gene therapy routes to treat osteoarthritis, stress and  
 CC apoptotic disorders. As such, they exhibit osteoprotective, antiarthritic and  
 CC tranquilizer activities. This oligonucleotide sequence is a PCR primer  
 CC used to construct the recombinant expression vector for human PRS51-L  
 CC cDNA of the invention.  
 XX  
 XX Sequence 34 BP; 4 A; 15 C; 8 G; 7 T; 0 U; 0 Other;  
 SQ  
 Query Match 2.0%; Score 21; DB 12; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1011 ATGCACCTGGCCCTCCGCC 1031  
 Db 14 ATGCACCTGGCCCTCCGCC 34  
 RESULT 7  
 ADG99925/c  
 ID ADG99925 standard; DNA; 244 BP.  
 XX  
 AC ADG99925;  
 XX  
 DT 11-MAR-2004 (first entry)  
 XX  
 DE Kidney disease-associated gene-related mouse DNA sequence Segid1055.  
 XX  
 KM Kidney disease; proximal tubule; nephrotropic; chronic renal failure;  
 KM diabetic nephropathy; mouse; murine; ds.  
 XX  
 OS Mus sp.  
 PN WO2003091427-A1.  
 PD 06-NOV-2003.  
 XX  
 PF 23-APR-2003; 2003WO-JP005137.  
 XX  
 PR 24-APR-2002; 2002JP-0012253.  
 XX  
 PA (KANS-) KANSAS TECHNOLOGY LICENSING ORG CO LTD.  
 XX  
 PI Takenaka M, Imai E, Okubo K;  
 DR WPI; 2004-011769/01.  
 XX  
 PT Kidney disease-associated genes expressed in proximal tubule, applicable

PT in diagnosis of, in drug development and in devising appropriate strategy  
 PT in treatment of e.g. chronic renal failure and diabetic nephropathy.  
 XX  
 XX Claim 1; SEQ ID NO 1055; 770pp; Japanese.  
 PS  
 CC This invention relates to novel kidney disease-associated genes expressed  
 CC in proximal tubules. The invention may be useful for the development of  
 CC compounds with a nephrotropic activity. The genes are applicable in the  
 CC diagnosis of, in drug development and in devising appropriate strategy in  
 CC the treatment of (for example) chronic renal failure and diabetic  
 CC nephropathy. The present sequence is that of a mouse DNA sequence which is  
 CC related to the invention.  
 XX  
 XX Sequence 244 BP; 55 A; 67 C; 65 G; 56 T; 0 U; 1 Other;  
 SQ  
 Query Match 2.0%; Score 21; DB 12; Length 244;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 903 TGTGGCTGAGGCTGGCTG 923  
 Db 102 TGTGGCTGAGGCTGGCTG 82  
 RESULT 8  
 ABL99890/c  
 ID ABL99890 standard; cDNA; 1795 BP.  
 XX  
 AC ABL99890;  
 XX  
 DT 03-OCT-2002 (first entry)  
 XX  
 DE Human secretory polynucleotide (sepm) 145.  
 XX  
 KM Human; ss; gene; secretory protein; secretory polynucleotides; SPTM;  
 KM SPTM-related disease; somatic gene therapy; germline gene therapy;  
 KM severe combined immunodeficiency; intracellular parasite protection;  
 KM fungal parasite; protozoan parasite; cell proliferative disorder; cancer;  
 KM immune disorder; AIDS; neurological disorder; Parkinson's disease;  
 KM motor neuron disorder; demyelinating disease; multiple sclerosis;  
 KM meningitis; abscess; prion diseases; cerebral palsy;  
 KM neuroskeletal disorder; peripheral nerve system disorder;  
 KM dermatomyositis; polymyositis; myopathy; myasthenia gravis;  
 KM mental disorder; Tourette's syndrome.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200220756-A2.  
 PD 14-MAR-2002.  
 XX  
 PF 30-AUG-2001; 2001WO-US027297.  
 XX  
 XX 05-SEP-2000; 2000US-0229747P.  
 PR 05-SEP-2000; 2000US-0229748P.  
 PR 05-SEP-2000; 2000US-0229749P.  
 PR 05-SEP-2000; 2000US-0229750P.  
 PR 05-SEP-2000; 2000US-0229751P.  
 PR 05-SEP-2000; 2000US-0230016P.  
 PR 05-SEP-2000; 2000US-0230583P.  
 PR 06-SEP-2000; 2000US-0230505P.  
 PR 06-SEP-2000; 2000US-0230514P.  
 PR 06-SEP-2000; 2000US-0230515P.  
 PR 06-SEP-2000; 2000US-0230517P.  
 PR 06-SEP-2000; 2000US-0230518P.  
 PR 06-SEP-2000; 2000US-0230519P.  
 PR 06-SEP-2000; 2000US-0230595P.  
 PR 06-SEP-2000; 2000US-0230596P.  
 PR 06-SEP-2000; 2000US-0230597P.  
 PR 06-SEP-2000; 2000US-0230599P.  
 PR 06-SEP-2000; 2000US-0230610P.  
 PR 06-SEP-2000; 2000US-0230664P.  
 PR 06-SEP-2000; 2000US-0230865P.

PR 06-SEP-2000; 2000US-0230988P.  
 PR 06-SEP-2000; 2000US-0230989P.  
 PR 06-SEP-2000; 2000US-0230990P.  
 PR 07-SEP-2000; 2000US-0230896P.  
 PR 07-SEP-2000; 2000US-0230897P.  
 PR 07-SEP-2000; 2000US-0230951P.  
 PR 07-SEP-2000; 2000US-0231163P.  
 PR 07-SEP-2000; 2000US-0231832P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JR,  
 PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
 PI Montiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
 PI Gerstein EH, Peralta CH, David MH, Panzer SR, Flores V, Dafio A,  
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;  
 DR WPI; 2002-315658/35.  
 XX P-PSDB; ABB97893.  
 XX  
 PT Polynucleotide sequences encoding human secretory proteins useful for  
 PT gene therapy of e.g. genetic deficiency disorders, cancers, and diseases  
 PT caused by intracellular parasites.  
 XX  
 PS Claim 1; Page 336-337; 585bp; English.  
 XX  
 CC The invention comprises the amino acid and coding sequences of human  
 CC secretory (SPTM) proteins. The SPTM DNA and amino acid sequences are  
 CC useful for creating a disease or condition associated with the expression  
 CC of functional SPTM. The SPTM DNA sequences are useful for somatic or  
 CC germline gene therapy to correct a genetic deficiency (e.g. severe  
 CC combined immunodeficiency). The SPTM DNA sequences are also useful in  
 CC providing protection against intracellular parasites (e.g. fungal  
 CC parasites and protozoan parasites). The SPTM DNA and protein sequences  
 CC are also useful for diagnosing cell proliferative disorders, cancer,  
 CC immune disorders (e.g. AIDS), neurological disorders (e.g. Parkinson's  
 CC disease), motor neuron disorders, demyelinating diseases (e.g. multiple  
 CC sclerosis), meningitis, abscesses, prion diseases, cerebral palsy,  
 CC neurodegenerative disorders, peripheral nervous system disorders,  
 CC dermatomyositis and polymyositis, myopathy, myasthenia gravis, and mental  
 CC disorders (e.g. Tourette's syndrome). cDNA sequences ABL99746 - ABL99929  
 CC represent human secretory polynucleotides of the invention  
 CC  
 XX  
 SQ Sequence 1795 BP; 350 A; 585 C; 502 G; 358 T; 0 U; 0 Other;  
 Query Match 2.0%; Score 21; DB 6; Length 1795;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1006 CACTGATGCACCTGACCTTC 1026  
 Db 624 CACTGATGCACCTGACCTTC 604  
 RESULT 9  
 ID ADQ87121/c  
 XX ADQ87121 standard; cDNA; 3157 BP.  
 AC ADQ87121;  
 XX  
 DT 07-OCT-2004 (first entry)  
 XX  
 DE Human tumour-associated antigenic target (TAT) cDNA sequence #3997.  
 XX  
 KW human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;  
 KW cancer; cell proliferative disorder; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W02004060270-A2.  
 XX  
 XX 22-JUL-2004.  
 PD  
 XX

PF 15-OCT-2003; 2003MO-US029126.  
 XX  
 XX  
 PR 18-OCT-2002; 2002US-0418988P.  
 XX  
 XX (GETH) GENENTECH INC.  
 PA (WUTD/) WU T D.  
 PA (ZHOU/) ZHOU Y.  
 XX  
 XX WU TD, Zhou Y;  
 DR WPI; 2004-534300/51.  
 XX  
 PT New nucleic acid molecule and encoded polypeptide, for diagnosing,  
 PT preventing or treating cell proliferative disorders such as cancer.  
 PS Claim 1; SEQ ID NO 3997; 5504bp; English.  
 XX  
 CC The present invention describes an isolated tumour-associated antigenic  
 CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide  
 CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of  
 CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%  
 CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-  
 CC (c). Also described: (1) an expression vector comprising the above  
 CC nucleic acid; (2) a host cell comprising the above expression vector; (3)  
 CC a process for producing a polypeptide; (4) an isolated polypeptide  
 CC comprising: (a) an amino acid sequence encoded by any of the above  
 CC nucleotide sequences; (b) an amino acid sequence encoded by the full-  
 CC length coding region of the above nucleotide sequences; or (c) a sequence  
 CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide  
 CC comprising the above polypeptide fused to a heterologous polypeptide; (6)  
 CC an isolated antibody that binds to the above polypeptide; (7) a process  
 CC for producing the antibody; (8) an isolated oligopeptide that binds to  
 CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)  
 CC binding organic molecule that binds to the above polypeptide; (10) a  
 CC composition of matter comprising the above (chimeric) polypeptide,  
 CC antibody, oligopeptide or TAT binding organic molecule, in combination  
 CC with a carrier; (11) an article of manufacture comprising a container and  
 CC the composition of matter contained within the container; (12) methods of  
 CC inhibiting the growth of a cell that expresses the above protein, where  
 CC the growth of the cell is at least in part dependent upon a growth  
 CC potentiating effect of the above protein; (13) a method of  
 CC therapeutically treating a mammal having a cancerous tumour comprising  
 CC cells that express the above protein; (14) a method of determining the  
 CC presence of a protein in a sample suspected of containing the protein  
 CC described above; (15) methods of diagnosing the presence of a tumour in a  
 CC mammal; (16) a method for treating or preventing a cell proliferative  
 CC disorder associated with increased expression or activity of the above  
 CC protein; and (17) a method of binding an antibody, oligopeptide or  
 CC organic molecule to a cell that expresses the protein described above.  
 CC The TAT sequences have cytostatic activities, and can be used in gene  
 CC therapy. The composition and methods are useful for diagnosing,  
 CC preventing or treating cancer. The composition is also used for preparing  
 CC a medicament for the therapeutic treatment or diagnostic detection of a  
 CC cell proliferative disorder or cancer. The present sequence represents a  
 CC human TAT cDNA sequence from the present invention.  
 CC  
 XX  
 SQ Sequence 3157 BP; 636 A; 1026 C; 812 G; 683 T; 0 U; 0 Other;  
 Query Match 2.0%; Score 21; DB 12; Length 3157;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1006 CACTGATGCACCTGACCTTC 1026  
 Db 541 CACTGATGCACCTGACCTTC 521  
 RESULT 10  
 ID AB236074/c  
 XX AB236074 standard; cDNA; 3983 BP.  
 AC AB236074;  
 XX

DT 10-FEB-2003 (first entry)  
 XX  
 DE Human secretory polynucleotide SPTM SEQ ID NO 238.  
 XX  
 KM Human, SPTM, autoimmune disorder; inflammatory disorder; AIDS; anaemia;  
 KM asthma; Crohn's disease; neurological disorder; epilepsy; cancer;  
 KM Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;  
 KM multiple sclerosis; Parkinson's disease; cell proliferative disorder;  
 KM anti-inflammatory; immunosuppressive; neuroprotective; nootropic;  
 KM neuroleptic; anticonvulsant; cyostatic; antiparkinsonian; anxiolytic;  
 KM antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;  
 KM secretory polynucleotide; secretory protein; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200283876-A2.  
 XX  
 PD 24-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-US009921.  
 XX  
 PR 29-MAR-2001; 2001US-028067P.  
 PR 29-MAR-2001; 2001US-028068P.  
 PR 16-MAY-2001; 2001US-0291280P.  
 PR 17-MAY-2001; 2001US-0291829P.  
 PR 17-MAY-2001; 2001US-0291849P.  
 PR 19-JUN-2001; 2001US-0289428P.  
 PR 20-JUN-2001; 2001US-029776P.  
 PR 20-JUN-2001; 2001US-0300001P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
 PI Dufour GE, Hillman JL, Yu JT, Tuason O, Yap PE, Amesley SR;  
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstein EH;  
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harrie B;  
 PI Flores V, Marwaha R, Lo A, Lan RY, Uraehka ME;  
 XX  
 DR WPI; 2003-075543/07.  
 DR P-PSDB; ABP75630.  
 XX  
 PT New human secretory proteins and polynucleotides, useful for diagnosing,  
 PT treating or preventing autoimmune/inflammatory disorders (e.g. AIDS),  
 PT neurological disorders (e.g. Alzheimer's), or cell proliferations or  
 PT cancers.  
 XX  
 PS Claim 1; SEQ ID NO 238; 458bp + Sequence Listing; English.  
 XX  
 CC The invention relates to a secretory polynucleotide (designated sptm)  
 CC comprising any of 567 polynucleotide sequences (ABZ35837-ABZ36403), a  
 CC naturally occurring polynucleotide sequence at least 90 % identical to  
 CC the polynucleotide sequence, a polynucleotide complementary to them or an  
 CC RNA equivalent of them. The polypeptide or polynucleotide are useful for  
 CC treating, preventing or diagnosing a disease or condition associated with  
 CC the expression of functional SPTM. These are particularly useful for  
 CC diagnosing, treating or preventing autoimmune/inflammatory disorders  
 CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's  
 CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,  
 CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,  
 CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,  
 CC schizophrenia or amnesia), or cell proliferative disorders (e.g.  
 CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,  
 CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,  
 CC breast, cervix or prostate). Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 3983 BP; 819 A; 1257 C; 980 G; 924 T; 0 U; 3 Other;

Cy 1006 CACTGATGCACCTGACCTTC 1026  
 Db 624 CACTGATGCACCTGACCTTC 604  
 XX  
 XX RESULT 11  
 XX ADM87103/c  
 XX ID ADM87103 standard; cDNA; 4030 BP.  
 XX  
 AC ADM87103;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Human protein encoding cDNA SEQ ID NO.196.  
 XX  
 KM respiratory; cyostatic; antiarthritic; antiinflammatory;  
 KM gastrointestinal; antibacterial; immunosuppressive; antidiabetic;  
 KM antirheumatic; gene therapy; molecular weight marker; chromosome marker;  
 KM chromosome tag; genetic fingerprinting; nutritional supplement; cancer;  
 KM inflammatory condition; arthritis; inflammatory bowel disease;  
 KM Crohn's disease; sepsis; rheumatoid arthritis; diabetes mellitus type 1;  
 KM graft versus host disease; human; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2004009834-A2.  
 XX  
 PD 29-JAN-2004.  
 XX  
 PF 19-JUL-2002; 2002WO-US022858.  
 XX  
 PR 21-JUL-2001; 2001US-0306971P.  
 PR 28-MAR-2002; 2002US-00112944.  
 XX  
 PA (NUVE-) NUVELO INC.  
 XX  
 PI Tang YT, Yang Y, Meng G, Zhang J, Ren F, Xue A, Wang J;  
 PI Wehrman T, Ghosh MJ, Wang D, Zhao QA, Wang Z;  
 XX  
 DR WPI; 2004-143291/14.  
 DR P-PSDB; ADM87347.  
 XX  
 PT New isolated polynucleotides and polypeptides, useful for treating, e.g.  
 PT cancer, lung or liver fibrosis, arthritis, inflammatory bowel disease,  
 PT Crohn's disease, rheumatoid arthritis, diabetes mellitus type 1 or graft  
 PT versus host disease.  
 XX  
 PS Claim 1; SEQ ID NO 196; 591bp; English.  
 XX  
 CC The present invention describes an isolated polynucleotide (I): (a)  
 CC comprising a nucleotide sequence selected from SEQ ID NO:1-244; or (b)  
 CC which encodes a polypeptide with biological activity, where the  
 CC polynucleotide hybridises to (I) under stringent hybridisation conditions  
 CC or has greater than 98% sequence identity with (I). (I) has respiratory,  
 CC cyostatic, antiarthritic, antiinflammatory, gastrointestinal,  
 CC antibacterial, immunosuppressive, antidiabetic and antirheumatic  
 CC activities, and can be used in gene therapy. (I) can be used for  
 CC generating polynucleotides encoding chimeric or fusion proteins and  
 CC heterologous protein sequences. The polynucleotides can be used to  
 CC express recombinant protein for analysis, characterisation or therapeutic  
 CC use; as markers for tissues in which the corresponding protein is  
 CC preferentially expressed; as molecular weight markers on gels; as  
 CC chromosome markers or tags to identify chromosomes or to map related gene  
 CC positions; to compare with endogenous DNA sequences in patients to  
 CC identify potential genetic disorders; as probes to hybridise and discover  
 CC genes, related DNA sequences; as a source of information to derive PCR  
 CC primers for genetic fingerprinting; as a probe to subtract-out known  
 CC sequences in the process of discovering other novel polynucleotides; for  
 CC selecting and making oligomers for attachment to a gene chip or other  
 CC support, including for examination of expression patterns; and as an antigen  
 CC to raise anti-DNA antibodies or elicit another immune response. The  
 CC polynucleotides and polypeptides can also be used as nutritional sources



CC or supplements, e.g. as a protein or amino acid supplement, as a carbon  
CC source, as a nitrogen source or as a source of carbohydrates. The  
CC polynucleotides and polypeptides can also be used to treat cancer. The  
CC compositions are useful for promoting better or faster closure of non-  
CC healing wounds, for the generation and regeneration of tissues, for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues, and conditions resulting from  
CC systemic cytokine damage. The compositions can also be used to treat  
CC inflammatory conditions (e.g. arthritis, inflammatory bowel disease or  
CC Crohn's disease), sepsis, rheumatoid arthritis, diabetes mellitus type 1  
CC or graft versus host disease. The present sequence represents a novel  
CC human polynucleotide sequence from the present invention. N.B. The  
CC sequences for this patent were obtained from the USPTO web site from an  
CC equivalent US patent US20040048249A1.  
XX  
SQ Sequence 4030 BP; 828 A; 1269 C; 1056 G; 876 T; 0 U; 1 Other;  
  
Query Match 2.0%; Score 21; DB 12; Length 4030;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1006 CACTGATGCACCTGGCCCTTC 1026  
DB 1508 CACTGATGCACCTGGCCCTTC 1488  
|||||  
  
RESULT 12  
ADQ80254  
ID ADQ80254 standard; cDNA; 127145 BP.  
XX  
AC ADQ80254;  
XX  
DT 21-OCT-2004 (first entry)  
XX  
DE Hermansky-Pudlak syndrome associated cDNA.  
XX  
XX 89; gene; cytostatic; epidermal growth factor receptor modulator;  
KM identification; therapeutic response; cancer; EGFR; biomarker.  
XX  
OS Homo sapiens.  
XX  
PN WO2004063709-A2.  
XX  
PD 29-JUL-2004.  
XX  
PF 08-JAN-2004; 2004WO-US000368.  
XX  
PR 08-JAN-2003; 2003US-0438735P.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Amier LC, Jannuario T;  
XX  
XX WPI; 2004-544114/52.  
DR P-PSDB; ADQ80373.  
XX  
PT Identifying a mammal that will respond therapeutically to a method of  
PT treating cancer comprises comparing the level of a biomarker in a mammal  
PT before and after exposure to an epidermal growth factor receptor (EGFR)  
PT modulator.  
XX  
XX  
PS Disclosure; SEQ ID NO 26; 520bp; English.  
XX  
XX The invention relates to a method of identifying a mammal that will  
CC respond therapeutically to a method of treating cancer by administering  
CC an epidermal growth factor receptor (EGFR) modulator by comparing the  
CC level of a biomarker in a mammal before and after exposure to an EGFR  
CC modulator. The method comprises: (a) measuring, in the mammal, the level  
CC of at least one biomarker identified in the specification; (b) exposing  
CC the mammal to the EGFR modulator; and (c) measuring in the mammal the  
CC level of the biomarker, where a difference in the level in step (c)  
CC compared to step (a) indicates that the mammal will respond  
CC therapeutically to the method of treating cancer. The method and

CC biomarkers are useful for identifying a mammal that will respond  
CC therapeutically to a method of treating cancer by administering an  
CC epidermal growth factor receptor (EGFR) modulator. This sequence  
CC corresponds to one of the biomarkers whose levels of gene expression is  
CC measured in the method of the invention.  
XX  
SQ Sequence 127145 BP; 31066 A; 31499 C; 32828 G; 31752 T; 0 U; 0 Other;  
  
Query Match 2.0%; Score 21; DB 13; Length 127145;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1006 CACTGATGCACCTGGCCCTTC 1026  
DB 58197 CACTGATGCACCTGGCCCTTC 58217  
|||||  
  
RESULT 13  
ACF62741/C  
ID ACF62741 standard; DNA; 189013 BP.  
XX  
AC ACF62741;  
XX  
DT 08-OCT-2003 (first entry)  
XX  
DE Cancer based on CYP3A5 related polynucleotide SEQ ID NO:669.  
XX  
XX Cancer; CYP3A5; irinotecan; pharmaceutical; malignant glioma;  
KM cytochrome p450; subfamily IIA; nifedipine oxidase; polypeptide 5;  
KW cytostatic; gene; ds.  
XX  
XX Unidentified.  
OS  
XX  
PN WO2003013534-A2.  
XX  
XX 20-FEB-2003.  
XX  
PD 23-JUL-2002; 2002WO-EP008219.  
XX  
PF 23-JUL-2001; 2001EP-00117608.  
XX  
PR 24-MAY-2002; 2002EP-00011710.  
XX  
XX (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.  
XX  
XX  
PI Heinrich G, Kerb R;  
XX  
XX WPI; 2003-268144/26.  
DR  
XX  
XX New use of irinotecan for preparation of compositions for treating cancer  
PT in subject having genome with variant allele comprising cytochrome p450,  
PT subfamily IIA, polypeptide 5 polynucleotide, termed CYP3A5.  
XX  
XX  
PS Disclosure; SEQ ID NO 669; 86bp; English.  
XX  
XX The present invention describes the use of irinotecan (I) or its  
CC derivative for the preparation of a pharmaceutical composition for  
CC treating colorectal, cervical, gastric, lung, ovarian or pancreatic  
CC cancer, or malignant glioma in a subject having a genome with a variant  
CC allele which comprises a cytochrome p450, subfamily IIA (nifedipine  
CC oxidase), polypeptide 5 (CYP3A5) polynucleotide (II). (I) is improved,  
CC cytostatic activity. The therapeutic applications of (I) is improved,  
CC since it is possible to individually treat a subject with an appropriate  
CC dosage and/or an appropriate derivative of (I). Therefore, undesirable,  
CC harmful or toxic effects are efficiently avoided. Unnecessary and  
CC potentially harmful treatment of those subjects who do not respond to the  
CC treatment with substances (nonresponders), as well as the development of  
CC drug resistances due to suboptimal drug dosing can be avoided. ACF62200  
CC to ACF62751 and ABM34912 to ABM35013 represent sequences used in the  
CC exemplification of the present invention  
XX  
SQ Sequence 189013 BP; 53603 A; 38975 C; 39348 G; 57087 T; 0 U; 0 Other;  
  
Query Match 2.0%; Score 21; DB 8; Length 189013;



Best Local Similarity 100.0%; Pred. No. 13;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 TGTCTGAGGCCACACAGCTAG 134  
DB 4162 TGTCTGAGGCCACACAGCTAG 4142

RESULT 14  
ADB20856/c  
ID ADB20856 standard; DNA; 189013 BP.

AC ADB20856;

XX 20-NOV-2003 (first entry)

DE MRP1 based cancer related nucleic acid SEQ ID NO:669.

XX irinotecan; colorectal cancer; cervical cancer; gastric cancer;  
KM lung cancer; ovarian cancer; pancreatic cancer; malignant glioma;  
KM variant allele; multidrug resistance protein 1; MRP1; cytosolic; gene;  
ds.

XX Unidentified.

PN WO2003013533-A2.

PD 20-FEB-2003.

PF 23-JUL-2002; 2002MO-EP008200.

XX 23-JUL-2001; 2001EP-00117608.

PR 24-MAY-2002; 2002EP-00011710.

XX (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.

PA Heinrich G, Korb R;

XX WPI; 2003-354397/33.

XX Use of irinotecan or its derivative for preparation of a pharmaceutical  
PT composition for treating cancer in a subject having a genome with a  
PT variant allele comprising a multidrug resistance protein 1  
PT polynucleotide.

XX Disclosure; SEQ ID NO 669; 100pp; English.

XX The present invention describes a method for the use of irinotecan (I) or  
CC its derivative for the preparation of a pharmaceutical composition for  
CC treating colorectal, cervical, gastric, lung, ovarian or pancreatic  
CC cancer, or malignant glioma in a subject having a genome with a variant  
CC allele which comprises a multidrug resistance protein 1 (MRP1)  
CC polynucleotide (II). (I) has cytostatic activity. (II) or its derivative  
CC can be used for the preparation of a pharmaceutical composition for  
CC treating colorectal, cervical, gastric, lung, ovarian or pancreatic  
CC cancer, or malignant glioma in a subject, where the subject is a human  
CC (preferably African or Asian) or a mouse. The present sequence represents  
CC a sequence which is used in the exemplification of the present invention.

XX SQ Sequence 189013 BP; 53603 A; 38975 C; 39348 G; 57087 T; 0 U; 0 Other;

Query Match 2.0%; Score 21; DB 8; Length 189013;

Best Local Similarity 100.0%; Pred. No. 13;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 TGTCTGAGGCCACACAGCTAG 134

DB 4162 TGTCTGAGGCCACACAGCTAG 4142

RESULT 15  
ADB87945/c  
ID ADB87945 standard; DNA; 189013 BP.

XX ADB87945;

XX 04-DEC-2003 (first entry)

DE Human UGT1A1 gene sequence SEQ ID NO:669.

XX irinotecan; cancer; UGT1A1; cytosolic; topoisomerase I inhibitor;  
KM colorectal cancer; cervical cancer; gastric cancer; lung cancer;  
KM ovarian cancer; pancreatic cancer; malignant glioma;  
KM uridine diphosphate glycosyltransferase1 member A1; gene; ds.

XX Homo sapiens.

PN WO2003013536-A2.

PD 20-FEB-2003.

PF 23-JUL-2002; 2002MO-EP008217.

XX 23-JUL-2001; 2001EP-00117608.

PR 24-MAY-2002; 2002EP-00011710.

XX (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.

PA Heinrich G, Korb R;

XX WPI; 2003-289896/28.

XX Use of irinotecan to treat cancer patient by determining if patient has  
PT variant alleles of UGT1A1 gene, administering increased/decreased amounts  
PT of irinotecan based on increased/decreased levels of UGT1A1 gene product.

XX Disclosure; SEQ ID NO 669; 107pp; English.

XX The invention relates to the novel use of irinotecan to treat a patient  
CC suffering from cancer. This involves determining if the patient has one  
CC or more variant alleles of the UGT1A1 gene, and if the patient has one or  
CC more of such variant alleles, irinotecan is administered in an increased  
CC or decreased amount in comparison to the amount that is administered  
CC without regard to the patient's alleles in the UGT1A1 gene. The invention  
CC has cytostatic activity. A composition of the invention acts as a  
CC topoisomerase I inhibitor. The method is useful for treating a patient,  
CC an animal e.g. mouse or a human, preferably African or Asian, suffering  
CC from cancer such as colorectal, cervical, gastric cancer, lung, ovarian,  
CC pancreatic cancer or malignant glioma. The present sequence is used in  
CC the exemplification of the invention.

XX SQ Sequence 189013 BP; 53603 A; 38975 C; 39348 G; 57087 T; 0 U; 0 Other;

Query Match 2.0%; Score 21; DB 10; Length 189013;

Best Local Similarity 100.0%; Pred. No. 13;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 TGTCTGAGGCCACACAGCTAG 134

DB 4162 TGTCTGAGGCCACACAGCTAG 4142

Search completed: June 29, 2005, 18:37:34  
Job time : 686 secs

**THIS PAGE BLANK (USPTO)**



```
Db      61 ATTGAGTCCATTTCAGATGAGAGTTGAGGCCCGAGAGAACTGATATCTGCTGA 120
Qy      121 GGCACACAGCTTGAAGAAGCAGCCAGCCGCAACCTCTGTGTGTGAGCCCGCAGC 180
Db      121 GGCACACAGCTTGAAGAAGCAGCCAGCCGCAACCTCTGTGTGTGAGCCCGCAGC 180
Qy      181 CCGATTGCTATTCGGGGGCTCGGGAGCCAGCCGAGGGTGTAGAGCATGTGTTCCAG 240
Db      181 CCGATTGCTATTCGGGGGCTCGGGAGCCAGCCGAGGGTGTAGAGCATGTGTTCCAG 240
Qy      241 TGGTGGGAATCTGAGAGAGCCCGGACAGGCCCCGTGAGGGAACCCCGAGGGCTGAGGC 300
Db      241 TGGTGGGAATCTGAGAGAGCCCGGACAGGCCCCGTGAGGGAACCCCGAGGGCTGAGGC 300
Qy      301 CCGGTGCTCATGATGCTCAGGCTGTGTGCTCTGAGCAGCCAGCCCTACTGTGACG 360
Db      301 CCGGTGCTCATGATGCTCAGGCTGTGTGCTCTGAGCAGCCAGCCCTACTGTGACG 360
Qy      361 GCAGCAGGAATCTGAGCCCGGGAAGGGTCCAGGGGAAGTTGTGAACCATCTAGCAAGTGG 420
Db      361 GCAGCAGGAATCTGAGCCCGGGAAGGGTCCAGGGGAAGTTGTGAACCATCTAGCAAGTGG 420
Qy      421 GGCCTGGGGTGTGGCCAACTTGAACAAGATGTAGGGCCCTGTGGAATTTGGCAG 480
Db      421 GGCCTGGGGTGTGGCCAACTTGAACAAGATGTAGGGCCCTGTGGAATTTGGCAG 480
Qy      481 CTCTTTTGGCCACAGAGGGGACGCTGTGTCTCGGGCTGGGTAGCTCAGAGGGTCACT 540
Db      481 CTCTTTTGGCCACAGAGGGGACGCTGTGTCTCGGGCTGGGTAGCTCAGAGGGTCACT 540
Qy      541 GGGGGTCTTCCATACACCCCGGCTGGGAAGTGTAGCCCGAGGGCTCGGAGGGAGC 600
Db      541 GGGGGTCTTCCATACACCCCGGCTGGGAAGTGTAGCCCGAGGGCTCGGAGGGAGC 600
Qy      601 AGCTGAGCCCATGAGAGAGAGGGCCAGTTCTCTCTGTAAAGGTAATGTGCTGAGATGAG 660
Db      601 AGCTGAGCCCATGAGAGAGAGGGCCAGTTCTCTCTGTAAAGGTAATGTGCTGAGATGAG 660
Qy      661 GGAACAGACAGAGCCCGAGGGGAACTAACCCGAGATCCAGCCCGGCTCACTCCGTGTG 720
Db      661 GGAACAGACAGAGCCCGAGGGGAACTAACCCGAGATCCAGCCCGGCTCACTCCGTGTG 720
Qy      721 GCTCAGGGGAAATCTCTCACTCTCTGTAGCCCTCGGAGCCAGCTAGAGGGGTCCAGTG 780
Db      721 GCTCAGGGGAAATCTCTCACTCTCTGTAGCCCTCGGAGCCAGCTAGAGGGGTCCAGTG 780
Qy      781 AGGGGGGTGAGGAAGCCAGCAGCTGGAAGCTTTTAACTCTCTGAGGAGCAGAGC 840
Db      781 AGGGGGGTGAGGAAGCCAGCAGCTGGAAGCTTTTAACTCTCTGAGGAGCAGAGC 840
Qy      841 CCGTTTCCAAATGCTGTGTGTCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
Db      841 CCGTTTCCAAATGCTGTGTGTCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
Qy      901 AGGTGGGGCTGAGAGCTGTGCTTGAACCTGGGGAAGAGGGGTCTCAGGAAGAGCTCTCTCC 960
Db      901 AGGTGGGGCTGAGAGCTGTGCTTGAACCTGGGGAAGAGGGGTCTCAGGAAGAGCTCTCTCC 960
Qy      961 TGGCCACTGGGCACTAGGCTCTGTGGAGCTGGCAGCATCTGATCTGATGACCTGG 1020
Db      961 TGGCCACTGGGCACTAGGCTCTGTGGAGCTGGCAGCATCTGATCTGATGACCTGG 1020
Qy      1021 CCGTTTCCCGGAGGAG 1038
Db      1021 CCGTTTCCCGGAGGAG 1038
```

RESULT 2  
US-10-617-443B-1  
; Sequence 1, Application US/10617443B  
; Publication No. US2005001977A1  
; GENERAL INFORMATION:

```
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Cailin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/617,443B
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patencin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3006
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-617-443B-1

Query Match      100.0%; Score 1038; DB 21; Length 3006;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1038; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CAGGAGCTCGAAGTTTGAGTCCCTCCACATCACTGTTCCACAGATGTGTAGAGGGCAT 60
Db      1 CAGGAGCTCGAAGTTTGAGTCCCTCCACATCACTGTTCCACAGATGTGTAGAGGGCAT 60
Qy      61 ATTGAGTCCATTTCAGATGAGAGTTGAGGCCAGAGAACTGAATCTGTCTGA 120
Db      61 ATTGAGTCCATTTCAGATGAGAGTTGAGGCCAGAGAACTGAATCTGTCTGA 120
Qy      121 GGCACACAGCTTGAAGAAGCAGCCAGCCGCAACCTCTGTGTGTGAGCCCGCAGC 180
Db      121 GGCACACAGCTTGAAGAAGCAGCCAGCCGCAACCTCTGTGTGTGAGCCCGCAGC 180
Qy      181 CCGATTGCTATTCGGGGGCTCGGGAGCCAGCCGAGGGTGTAGAGCATGTGTTCCAG 240
Db      181 CCGATTGCTATTCGGGGGCTCGGGAGCCAGCCGAGGGTGTAGAGCATGTGTTCCAG 240
Qy      241 TGGTGGGAATCTGAGAGAGCCCGGACAGGCCCTGTGAGGGAACCCCGAGGGCTGAGGC 300
Db      241 TGGTGGGAATCTGAGAGAGCCCGGACAGGCCCTGTGAGGGAACCCCGAGGGCTGAGGC 300
Qy      301 CCGGTGCTCATGATGCTCAGGCTGTGTGTCTGTGAGCCAGCCCTCACTGCTGAGC 360
Db      301 CCGGTGCTCATGATGCTCAGGCTGTGTGTCTGTGAGCCAGCCCTCACTGCTGAGC 360
Qy      361 GGAAGAGAAATCTGAGCCCGGGAAGGGTCCAGGGAAGTTGTGAACCTAGCAAGTGG 420
Db      361 GGAAGAGAAATCTGAGCCCGGGAAGGGTCCAGGGAAGTTGTGAACCTAGCAAGTGG 420
Qy      421 GGCCTGGGGTGTGGCCAACTTGAACAAGATGTAGGGCCCTGTGGAATTTGGCAG 480
Db      421 GGCCTGGGGTGTGGCCAACTTGAACAAGATGTAGGGCCCTGTGGAATTTGGCAG 480
Qy      481 CTCTTTTGGCCACAGAGGGGACGCTGTGTCTCGGGCTGGGTAGCTCAGAGGGTCACT 540
Db      481 CTCTTTTGGCCACAGAGGGGACGCTGTGTCTCGGGCTGGGTAGCTCAGAGGGTCACT 540
Qy      541 GGGGGTCTTCCATACACCCCGGCTGGGAAGTGTAGCCCGAGGGCTCGGAGGGAGC 600
Db      541 GGGGGTCTTCCATACACCCCGGCTGGGAAGTGTAGCCCGAGGGCTCGGAGGGAGC 600
Qy      601 AGCTGAGCCCATGAGAGAGGGCCAGTTCTCTCTGTAAAGGTAATGTGATGAGCAG 660
Db      601 AGCTGAGCCCATGAGAGAGGGCCAGTTCTCTCTGTAAAGGTAATGTGATGAGCAG 660
Qy      661 GGAACAGACAGAGCCCGAGGGGAACTAACCCGAGATCCAGCCCGGCTCACTCCGTGTG 720
Db      661 GGAACAGACAGAGCCCGAGGGGAACTAACCCGAGATCCAGCCCGGCTCACTCCGTGTG 720
Qy      721 GCTCAGGGGAAATCTCTCACTCTCTGTAGCCCTCGGAGCCAGCTAGAGGGGTCCAGTG 780
Db      721 GCTCAGGGGAAATCTCTCACTCTCTGTAGCCCTCGGAGCCAGCTAGAGGGGTCCAGTG 780
Qy      781 AGGGGGGTGAGGAAGCCAGCAGCTGGAAGCTTTTAACTCTCTGAGGAGCAGAGC 840
```



```
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 5
/ LENGTH: 34
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: primer
US-10-189-099A-5

Query Match
Best Local Similarity 100.0%; Score 21; DB 17; Length 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1011 ATGCACCTGGCCCTTCCGCC 1031
Db 14 ATGCACCTGGCCCTTCCGCC 34

RESULT 7
US-10-617-443B-5
/ Sequence 5, Application US/10617443B
/ Publication No. US20050019777A1
/ GENERAL INFORMATION:
/ APPLICANT: Darrow, Andrew L
/ APPLICANT: Qi, Jian-shen
/ APPLICANT: Chen, Caixin
/ APPLICANT: Andrade-Gordon, Patricia
/ TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
/ FILE REFERENCE: CRT-1644
/ CURRENT APPLICATION NUMBER: US/10/617,443B
/ CURRENT FILING DATE: 2003-07-02
/ NUMBER OF SEQ ID NOS: 8
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 5
/ LENGTH: 34
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: primer
US-10-617-443B-5

Query Match
Best Local Similarity 100.0%; Score 21; DB 21; Length 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1011 ATGCACCTGGCCCTTCCGCC 1031
Db 14 ATGCACCTGGCCCTTCCGCC 34

RESULT 8
US-10-112-944-196/c
/ Sequence 196, Application US/10112944
/ Publication No. US2004008249A1
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Yang, Yonghong
/ APPLICANT: Wang, Gezhi
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyun
/ APPLICANT: Xue, Aidong J.
/ APPLICANT: Wang, Jian-Rui
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Ghosh, Malabika
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Wang, Zhiwei
/ TITLE OF INVENTION: No. US2004008249A1el Nucleic Acids and
/ FILE REFERENCE: 805A
/ CURRENT APPLICATION NUMBER: US/10/112,944
/ CURRENT FILING DATE: 2002-03-28
/ PRIOR APPLICATION NUMBER: US 09/488,725
```

```
/ PRIOR FILING DATE: 2000-01-21
/ PRIOR APPLICATION NUMBER: US 09/491,404
/ PRIOR FILING DATE: 2000-01-25
/ PRIOR APPLICATION NUMBER: US 09/496,914
/ PRIOR FILING DATE: 2000-02-03
/ PRIOR APPLICATION NUMBER: US 09/515,126
/ PRIOR FILING DATE: 2000-02-28
/ PRIOR APPLICATION NUMBER: US 09/519,705
/ PRIOR FILING DATE: 2000-03-07
/ PRIOR APPLICATION NUMBER: US 09/540,217
/ PRIOR FILING DATE: 2000-03-31
/ PRIOR APPLICATION NUMBER: US 09/552,929
/ PRIOR FILING DATE: 2000-04-18
/ PRIOR APPLICATION NUMBER: US 09/577,408
/ PRIOR FILING DATE: 2000-05-18
/ NUMBER OF SEQ ID NOS: 924
/ SOFTWARE: pc FL_genes Version 5.0
/ SEQ ID NO 196
/ LENGTH: 4030
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (985)..(3036)
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(4030)
/ OTHER INFORMATION: n = a,t,c or g
US-10-112-944-196

Query Match
Best Local Similarity 100.0%; Score 21; DB 18; Length 4030;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1006 CACTGATGCACCTGGCCCTTC 1026
Db 1508 CACTGATGCACCTGGCCCTTC 1488

RESULT 9
US-10-484-577-669/c
/ Sequence 669, Application US/10484577
/ Publication No. US20050032724A1
/ GENERAL INFORMATION:
/ APPLICANT: EPIDAUROS Biotechnologie Aktiengesellschaft
/ TITLE OF INVENTION: Means and methods for improved treatment of cancer based on UGT1A1
/ FILE REFERENCE: F2285PCT-1
/ CURRENT APPLICATION NUMBER: US/10/484,577
/ PRIOR FILING DATE: 2004-01-22
/ PRIOR APPLICATION NUMBER: PCT/EP 02/08220
/ PRIOR FILING DATE: 2002-07-23
/ PRIOR APPLICATION NUMBER: EP 01 11 7608.8
/ PRIOR FILING DATE: 2001-07-23
/ PRIOR APPLICATION NUMBER: EP 02011710.7
/ PRIOR FILING DATE: 2002-05-24
/ NUMBER OF SEQ ID NOS: 683
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 669
/ LENGTH: 189013
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-484-577-669

Query Match
Best Local Similarity 100.0%; Score 21; DB 21; Length 189013;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 TGCTGAGGCCACACGACTAG 134
Db 4162 TGCTGAGGCCACACGACTAG 4142

RESULT 10
```

```

US-10-189-099A-7
; Sequence 7, Application US/10189099A
; Publication No. US20040005659A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Callin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/189,099A
; CURRENT FILING DATE: 2002-12-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Internal Sequence
; FEATURE:
; OTHER INFORMATION: Description of Internal Sequence: primer
US-10-189-099A-7

```

```

Query Match 1.9%; Score 20; DB 17; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 413 GCAAGTCGGGCTGGGCTGTG 432
Db 1 GCAAGTCGGGCTGGGCTGTG 20

```

```

RESULT 11
US-10-617-443B-7
; Sequence 7, Application US/10617443B
; Publication No. US2005001977A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Callin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/617,443B
; CURRENT FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Internal primer
US-10-617-443B-7

```

```

Query Match 1.9%; Score 20; DB 21; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 413 GCAAGTCGGGCTGGGCTGTG 432
Db 1 GCAAGTCGGGCTGGGCTGTG 20

```

```

RESULT 12
US-10-027-632-195771/c
; Sequence 195771, Application US/10027632
; Publication No. US2002019837A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

```

```

; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195771
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-195771

```

```

Query Match 1.3%; Score 20; DB 13; Length 504;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 67 TCCCATTTTCAGATGAGA 86
Db 464 TCCCATTTTCAGATGAGA 445

```

```

RESULT 13
US-10-027-632-195771/c
; Sequence 195771, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195771
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-195771

```

```

Query Match 1.9%; Score 20; DB 17; Length 504;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 67 TCCCATTTTCAGATGAGA 86
Db 464 TCCCATTTTCAGATGAGA 445

```

RESULT 14

US-09-922-217-1058/C  
 ; Sequence 1058, Application US/09922217  
 ; Patent No. US200207641A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Meagher, Madeleine Joy  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Wang, Tongtong  
 ; APPLICANT: Jiang, Yugu  
 ; APPLICANT: Smith, Carole Lynn  
 ; APPLICANT: King, Gordon E.  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Clapper, Jonathan D.  
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
 ; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
 ; FILE REFERENCE: 210121.471C13  
 ; CURRENT APPLICATION NUMBER: US/09/922,217  
 ; NUMBER OF SEQ ID NOS: 1124  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1058  
 ; LENGTH: 15720  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-922-217-1058

Query Match 1.9%; Score 20; DB 9; Length 15720;  
 Best Local Similarity 100.0%; Pred. No. 5.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 316 GCCTCAGGCTGTGTCTG 335  
 DB 13006 GCCTCAGGCTGTGTCTG 12987

RESULT 15  
 US-09-833-263-1058/C  
 ; Sequence 1058, Application US/09833263  
 ; Patent No. US20020110547A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Clapper, Jonathan D.  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Meagher, Madeleine J.  
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE  
 ; FILE REFERENCE: 210121.471C12  
 ; CURRENT APPLICATION NUMBER: US/09/833,263  
 ; CURRENT FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 1093  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 1058  
 ; LENGTH: 15720  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-833-263-1058

Query Match 1.9%; Score 20; DB 9; Length 15720;  
 Best Local Similarity 100.0%; Pred. No. 5.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 316 GCCTCAGGCTGTGTCTG 335  
 DB 13006 GCCTCAGGCTGTGTCTG 12987

Search completed: June 29, 2005, 21:23:40  
 Job time : 763 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2005, 17:01:40 ; Search time 4886 Seconds  
(without alignments)  
10294.006 Million cell updates/sec

Title: US-10-617-443b-1\_COPY\_1\_1038  
Perfect score: 1038  
Sequence: 1 caggagactcgaagcttcgacg.....ggccctcccgccgacgacgacg 1038

Scoring table: Oligo\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4708233 seqs, 24227607955 residues

Word-size-1-12

Total number of hits satisfying chosen parameters: 730606

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmb1:\*  
1: gb ba:\*  
2: gb bta:\*  
3: gb in:\*  
4: gb om:\*  
5: gb ov:\*  
6: gb pat:\*  
7: gb ph:\*  
8: gb pl:\*  
9: gb pr:\*  
10: gb ro:\*  
11: gb sts:\*  
12: gb sy:\*  
13: gb un:\*  
14: gb vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	834	80.3	60597	9 AC113611	AC113611 Homo sapi
2	26	2.5	235102	2 AC118993	AC118993 Rattus no
3	26	2.5	238368	2 AC110369	AC110369 Rattus no
4	26	2.5	264075	2 AC134075	AC134075 Rattus no
5	24	2.3	145023	9 AC012621	AC012621 Homo sapi
6	24	2.3	164564	9 HUAC003108	AC003108 Human Chr
7	22	2.1	86027	9 AC137767	AC137767 Homo sapi
8	22	2.1	91841	9 AL831755	AL831755 Human DNA
9	22	2.1	97559	9 AL356748	AL356748 Human DNA
10	22	2.1	106497	9 AL157934	AL157934 Human DNA
11	22	2.1	112453	6 C0869788	C0869788 Sequence
12	22	2.1	125836	2 AC137629	AC137629 Homo sapi
13	22	2.1	128440	2 AC0002419	AC0002419 Homo sapi
14	22	2.1	131184	10 AL844513	AL844513 Mouse DNA
15	22	2.1	132118	9 AC104081	AC104081 Homo sapi
16	22	2.1	180396	9 AL445683	AL445683 Human DNA
17	22	2.1	183085	9 AC005815	AC005815 Homo sapi
18	22	2.1	192391	9 AC010768	AC010768 Homo sapi
19	22	2.1	193390	10 AC101527	AC101527 Mus muscu

c	20	22	2.1	197549	2 AC145423	AC145423 Homo sapi
c	21	22	2.1	220232	8 AC006288	AC006288 Homo sapi
c	22	22	2.1	233232	10 AL845161	AL845161 Mouse DNA
c	23	22	2.1	233874	2 BX842698	BX842698 Mus muscu
c	24	22	2.1	237538	2 AC108561	AC108561 Rattus no
c	25	22	2.1	243195	2 AC094523	AC094523 Rattus no
c	26	22	2.1	251121	2 AC128818	AC128818 Rattus no
c	27	21	2.0	1377	10 AF172275	AF172275 Mus muscu
c	28	21	2.0	1680	10 BC083769	BC083769 Rattus no
c	29	21	2.0	1795	6 AX740556	AX740556 Sequence
c	30	21	2.0	2045	10 BC026545	BC026545 Mus muscu
c	31	21	2.0	2055	6 CQ730199	CQ730199 Sequence
c	32	21	2.0	2184	9 CR456444	CR456444 Homo sapi
c	33	21	2.0	3088	9 AK025714	AK025714 Homo sapi
c	34	21	2.0	3200	9 BC015726	BC015726 Homo sapi
c	35	21	2.0	3236	9 BC040351	BC040351 Homo sapi
c	36	21	2.0	3944	9 BC027716	BC027716 Homo sapi
c	37	21	2.0	6112	10 AF075576	AF075576 Mus muscu
c	38	21	2.0	13178	9 AY504960	AY504960 Homo sapi
c	39	21	2.0	14268	10 AF338323	AF338323 Mus muscu
c	40	21	2.0	87077	9 HUAC004097	AC004097 Human Chr
c	41	21	2.0	91639	9 AL162399	AL162399 Human DNA
c	42	21	2.0	95283	9 HS599F21	AL035662 Human DNA
c	43	21	2.0	110000	2 AC106346_1	Continuation (2 of
c	44	21	2.0	112071	4 BX649374	BX649374 Smtithops
c	45	21	2.0	120873	9 AL445986	AL445986 Human DNA

# ALIGNMENTS

RESULT 1	AC113611	60597 bp	DNA	linear	PRI 10-MAR-2003
LOCUS	AC113611				
DEFINITION	Homo sapiens BAC clone RP11-421M20 from 4, complete sequence.				
ACCESSION	AC113611				
VERSION	AC113611.3	GI:28475761			
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Isak, A. and Cotton, M.				
TITLE	The sequence of Homo sapiens BAC clone RP11-421M20				
JOURNAL	Unpublished (2001)				
ACCESSION	3 (bases 1 to 60597)				
PUBMED	99063792				
REFERENCE	2 (bases 1 to 60597)				
AUTHORS	Isak, A. and Cotton, M.				
TITLE	The sequence of Homo sapiens BAC clone RP11-421M20				
JOURNAL	Unpublished (2001)				
REFERENCE	3 (bases 1 to 60597)				
AUTHORS	Waterston, R.H.				
TITLE	Submitted (04-MAR-2002) Genome Sequencing Center, Washington				
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	4 (bases 1 to 60597)				
AUTHORS	Waterston, R.H.				
TITLE	Submitted (11-APR-2002) Genome Sequencing Center, Washington				
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	5 (bases 1 to 60597)				
AUTHORS	Waterston, R.H.				
TITLE	Submitted (23-FEB-2003) Genome Sequencing Center, Washington				
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	6 (bases 1 to 60597)				
AUTHORS	Waterston, R.				
TITLE	Direct Submission				

JOURNAL  
Submitted (10-MAR-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Feb 23, 2003 this sequence version replaced gi:20136967.

## ----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>Contact: [saplens@wustl.edu](mailto:saplens@wustl.edu)

----- Summary Statistics

Center project name: H\_NH0421M20

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

## ----- MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see <http://genome.wustl.edu/gsc>

## ----- SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male  
donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E.,  
Tateno, M., Caranese, J.J. and de Jong, P.J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong  
and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

## ----- NEIGHBORING SEQUENCE INFORMATION:

This sequence is not the entire insert of the clone. This clone is  
overlapped by AC104650 and AC104825.

Discrepant bases between AC104650, AC104825 and clone sequence.

Data from AC104825 was used to finish this clone.

## FEATURES

## Source

Location/Qualifiers

1. .60597  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-421M20"  
/clone\_1id="RPC1-11"  
539. .903  
/rpt\_family="ERV.L"  
2797. .2907  
/rpt\_family="CAGAGA)n"  
2908. .3114  
/rpt\_family="ERV.L"  
3122. .3184  
/rpt\_family="CAGAGA)n"  
3256. .3368  
/rpt\_family="L1"  
3500. .3682  
/rpt\_family="L1"  
3713. .3782  
/rpt\_family="AT\_rich"  
5208. .5410  
/rpt\_family="MIR"

repeat\_region 5451. .5503  
/rpt\_family="MIR"  
repeat\_region 5504. .5677  
/rpt\_family="MIR"  
repeat\_region 6645. .6687  
/rpt\_family="MIR"  
repeat\_region 6998. .7067  
/rpt\_family="MIR"  
repeat\_region 7741. .8173  
/rpt\_family="MER1\_type"  
8530. .8610  
/rpt\_family="MIR"  
misc\_feature 8638. .9514  
/note="CpG island (%C=76.7, o/e=0.80, #CpGs=114)"  
repeat\_region 9027. .9048  
/rpt\_family="GC\_rich"  
repeat\_region 9063. .9111  
/rpt\_family="GC\_rich"  
repeat\_region 9612. .9739  
/rpt\_family="MIR"  
repeat\_region 10414. .10509  
/rpt\_family="MER103"  
10688. .11034  
/rpt\_family="MALR"  
11761. .11997  
/rpt\_family="L2"  
12212. .12325  
/rpt\_family="MIR"  
repeat\_region 13275. .13625  
/rpt\_family="L2"  
13703. .13769  
/rpt\_family="CR1"  
14926. .15054  
/rpt\_family="MIR"  
15239. .15281  
/rpt\_family="tip100"  
15349. .15515  
/rpt\_family="MIR"  
16000. .16113  
/rpt\_family="MIR"  
repeat\_region 16639. .16917  
/rpt\_family="L2"  
17127. .17228  
/rpt\_family="tip100"  
17883. .17982  
/rpt\_family="MIR"  
repeat\_region 18108. .18173  
/rpt\_family="CT\_rich"  
19059. .19127  
/rpt\_family="L2"  
20087. .20234  
/rpt\_family="MIR"  
20576. .20738  
/rpt\_family="MIR"  
22242. .22380  
/rpt\_family="TGA)n"  
23733. .23804  
/rpt\_family="L2"  
23890. .23987  
/rpt\_family="MIR"  
24011. .24440  
/rpt\_family="MALR"  
25254. .25412  
/rpt\_family="MIR"  
26001. .26150  
/rpt\_family="MIR"  
26473. .26512  
/rpt\_family="MIR"  
26513. .26894  
/rpt\_family="MALR"  
26895. .27048  
/rpt\_family="MIR"  
repeat\_region 27423. .27724

1 (Bases 1 to 235102)

Muzny, D., Matzie, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angilino, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnesstead, M., Benhamed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dedert, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinth, H., Divya, K., Draper, H., Durson-Socha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Faller, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Gamba, R., Garcia, A., Garner, T., Garza, M., Gebregergs, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, C., Kowis, C., Kraft, C. L., Lebow, H., Lervan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lounsburow, L., Lounsged, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Nandasa, M., Murphy, M., Nair, L., Naranvatsy, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelmele, O., Okunolu, G., Olarinmusegun, A., Pal, S., Parks, K., Pasternack, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L., -L., Puzos, M., Quiroz, J., Racilin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodney, T., Rojan, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, M., Savery, G., Scheer, S., Scott, G., Shatusman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sodergren, E., Song, X., -Z. Sorelle, R., Soes, J.,

TITLE	JOURNAL
Steinle,M., Strong,R., Sutton,A., Swack,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K., Valae,R., Vera,V., Villaena,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlezysk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yi,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.	Direct Submission Unpublished 2 (bases 1 to 235102) Morley,K.C. Direct Submission Submitted (24-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 235102) Rat Genome Sequencing Consortium. Direct Submission Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Nov 15, 2002 this sequence version replaced gi:23681105. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('a' contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: hgsc-help@bcm.tmc.edu Project information Center project name: GTNH Center clone name: CH230-237H8 Summary Statistics Assembly program: Phrap; version 0.990329 Consensus quality: 227037 bases at least Q40 Consensus quality: 229537 bases at least Q30 Consensus quality: 231225 bases at least Q20 Baitmated insert size: 235071; sum-of-contigs estimation Quality coverage: 8x in Q20 bases; sum-of-contigs estimation	----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: hgsc-help@bcm.tmc.edu Project information Center project name: GTNH Center clone name: CH230-237H8 Summary Statistics Assembly program: Phrap; version 0.990329 Consensus quality: 227037 bases at least Q40 Consensus quality: 229537 bases at least Q30 Consensus quality: 231225 bases at least Q20 Baitmated insert size: 235071; sum-of-contigs estimation Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
* NOTE: Baitmated insert size may differ from sequence length * (see <a href="http://www.hgsc.bcm.tmc.edu/docs/genbank_drafc_data.html">http://www.hgsc.bcm.tmc.edu/docs/genbank_drafc_data.html</a> ) * NOTE: This is a 'working draft' sequence. It currently * consists of 2 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. 1 108068: contig of 108068 bp in length * 108069 108168: gap of unknown length * 108169 235102: contig of 126934 bp in length. Location/Qualifiers 1..235102 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-237H8" 1..1143 /note="wgs_end_extension"	----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: hgsc-help@bcm.tmc.edu Project information Center project name: GTNH Center clone name: CH230-237H8 Summary Statistics Assembly program: Phrap; version 0.990329 Consensus quality: 227037 bases at least Q40 Consensus quality: 229537 bases at least Q30 Consensus quality: 231225 bases at least Q20 Baitmated insert size: 235071; sum-of-contigs estimation Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
misc_feature	

[illegible]

Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Unpublished  
2 (bases 1 to 258368)  
Worley, K. C.

Direct Submission  
Submitted (11-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 258368)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 19, 2002 this sequence version replaced gi:21744398.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atl/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: GRPY  
Center clone name: CH230-49K10

----- Summary Statistics -----  
Assembly program: Phred; version 0.990129  
Consensus quality: 236348 bases at least Q40  
Consensus quality: 236877 bases at least Q30  
Consensus quality: 240713 bases at least Q20  
Estimated insert size: 259061; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbankdraft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html)) -----  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 102159: contig of 102159 bp in length  
\* 102260: gap of unknown length  
\* 102260 249257: contig of 146998 bp in length  
\* 249258 249358: gap of unknown length  
\* 249358 258368: contig of 9011 bp in length.

Location/Qualifiers  
1. 258368  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-49K10"  
1. 1248  
/note="wgs\_end\_extension  
clone\_end:17"  
6211..7032

misc\_feature  
misc\_feature  
misc\_feature

Query Match 2.5%; Score 26; DB 2; Length 258368;  
Best Local Similarity 100.0%; Pred. No. 0.0057;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN  
misc\_feature  
misc\_feature  
misc\_feature  
/note="wgs\_contig"  
102260..104120  
/note="wgs\_contig"

Db 82091 CATGCCACGACCTGTCCTGCA 82066

ORIGIN  
AC134075  
LOCUS  
DEFINITION  
AC134075 264075 bp DNA linear HTG 13-NOV-2002  
Rattus norvegicus clone CH230-2H18, WORKING DRAFT SEQUENCE, 2  
unordered pieces.  
AC134075  
AC134075.2 GI:24941374  
HTG; HTS PHASE1; HTS\_DRAFT; HTS\_FULLTOP.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS  
1 (bases 1 to 264075)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Alt-omman, F.R., Allen, C., Alsbrooks, S.L., Amaralunga, H.C., Are, U.R., Ayele, M., Banks, T., Barbieria, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Bhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delataye, K.R., Delgado, O., Dunn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Fails, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homai, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karleson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kravovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, L., Louie, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, R., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabadi, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Oguh, M., Okunolu, G., Oragunye, N., Oriado, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Fu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojudo, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshani, N., Slason, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tameris, A., Tameris, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vaquez, L., Vera, V., Villalana, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 264075)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (22-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 264075)  
Worley,K.C.  
REFERENCE Direct Submission  
AUTHORS Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
COMMENT On Nov 13, 2002 this sequence version replaced gi:23267440. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: TUG2  
Center clone name: CH230-2H18  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 225336 bases at least Q40  
Consensus quality: 228408 bases at least Q30  
Consensus quality: 230612 bases at least Q20  
Estimated insert size: 229980; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 262529: contig of 262529 bp in length  
\* 262530 262629: gap of unknown length  
\* 262630 264075: contig of 1446 bp in length.  
Location/Qualifiers  
1. 264075  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-2H18"  
1. 1475  
/note="wge\_end\_extension  
clone\_end:T7  
1858. 153329  
/note="clone\_boundary  
clone\_end:T7  
site:  
end\_sequence:BH287375"  
258169. 258875  
/note="clone\_boundary  
clone\_end:Sp6

misc\_feature  
misc\_feature  
misc\_feature

misc\_feature site:  
end\_sequence:BH287377"  
259534. 262529  
/note="wge\_end\_extension  
clone\_end:Sp6"

ORIGIN  
Query Match 2.5%; Score 26; DB 2; Length 264075;  
Best Local Similarity 100.0%; Pred. No. 0.0057;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 313 CATGCTCAGGCTGTGTCTGTGCA 338  
DB 88708 CATGCTCAGGCTGTGTGTGTGCA 88733

RESULT 5  
AC012621 145023 bp DNA linear PRI 18-MAR-2003  
LOCUS Homo sapiens chromosome 16 clone CTD-2055J6, complete sequence.  
DEFINITION AC012621  
ACCESSION AC012621  
VERSION AC012621.8 GI:29029223  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 145023)  
DOE Joint Genome Institute, Stanford Human Genome Center and Los  
Alamos National Laboratory.  
Direct Submission  
2 (bases 1 to 145023)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 145023)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (06-FEB-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 145023)  
DOE Joint Genome Institute, Stanford Human Genome Center and Los  
Alamos National Laboratory.  
Direct Submission  
Submitted (18-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Mar 18, 2003 this sequence version replaced gi:18542968.  
Draft Sequence Produced by DOE Joint Genome Institute  
[www.jgi.doe.gov](http://www.jgi.doe.gov)  
Finishing Completed at Stanford Human Genome Center and Los Alamos  
National Laboratory  
[www-hgsc.stanford.edu](http://www-hgsc.stanford.edu)  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.  
Location/Qualifiers  
1. 145023  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="CTD-2055J6"

FEATURES  
source  
1. 145023  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="CTD-2055J6"

ORIGIN  
Query Match 2.3%; Score 24; DB 9; Length 145023;  
Best Local Similarity 100.0%; Pred. No. 0.085;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 TGAGGCCACACAGCTAGAAAGCAG 141  
DB 99318 TGAGGCCACACAGCTAGAAAGCAG 99341

RESULT 6  
HUC003108  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
SOURCE

HUC003108 164564 bp DNA linear PRI 30-OCT-2002  
Human Chromosome 16 BAC clone C1987SK-327024, complete sequence.  
AC003108  
AC003108.1 GI:2833632  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 164564)  
Loftus,B.J., Kim,U.J., Sneddon,V.P., Brandon,R.,  
Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,  
Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,  
Eichler,E.B., Harris,P.C., Venter,J.C. and Adams,M.D.  
Genome duplications and other features in 12 Mb of DNA sequence  
from human chromosome 16p and 16q  
Genomics 60 (3), 295-308 (1999)  
10493829  
2 (bases 1 to 164564)  
Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,  
Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.  
Human Chromosome 16 BAC clone C1987SK-327024  
Unpublished  
3 (bases 1 to 164564)  
Adams,M.D. and Loftus,B.J.  
Direct Submission  
Submitted (19-NOV-1997) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA, Email:  
bjoftus@tigr.org  
4 (bases 1 to 164564)  
Adams,M.D.  
Direct Submission  
Submitted (05-FEB-1998) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
5 (bases 1 to 164564)  
Adams,M.D. and Loftus,B.J.  
Direct Submission  
Submitted (17-MAR-1998) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
On Feb 5, 1998 this sequence version replaced gi:2827780.  
Address all correspondence to: Mark Adams The Institute for Genomic  
Research 9712 Medical Center Dr., Rockville, MD 20850, USA e-mail  
address: mdams@tigr.org. The orientation of the sequence is from  
SP6 end to T7 end. Genes were identified by a combination of five  
methods including: XGRAIL (available by anonymous ftp from  
archur.epm.ornl.gov), GeneFinder (Phil Green, University of  
Washington), GenScan (Chris Burge,  
http://genomic.stanford.edu/~chris/BURGE/GENSCAN.html) searches of the  
complete sequence against a peptide database, and the Human gene  
index database at TIGR (http://www.tigr.org/tdb/hgi.html).  
Genes without peptide homology having applied EST hits are termed  
'unknown gene product'. Genes encoding tRNAs are predicted by  
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).  
Location/Qualifiers  
1..164564  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16p11.2-12"  
/clone="327024"  
9637..9767  
/db\_xref="dbSTS:G21493"  
complement(10088..28580)  
/gene="327024.1"  
/note="327024.1"  
complement(join(10088..10235,11499..11710,14305..14397,  
17457..17562,23632..23807,28068..28253,28320..28580))  
/gene="327024.1"

CDS  
misc\_feature  
gene  
mRNA  
CDS

complement(join(10088..10235,11499..11710,14305..14397,  
17457..17562,23632..23807,28068..28253,28320..28580))  
/gene="327024.1"  
/codon\_start=1  
/product="unknown gene product"  
/protein\_id="AAC05803.1"  
/translation="MMTMDQGLGPFSLTLLVLTSTRSPVNAACLTGSLFVLTAV  
FSEPPVPSGALGVLRKRLISAIARHGSHDAPENTLAIROGSPFISGRPSHFLR  
PGCTPALDPSGGLGSLGVLYOKTADMYVTLTTLTAIRNNAAGATGVL  
DIEFTSGITVLTMDNTVDRTTGTGRLCLTTEQIRKLPANRRLANDPPEKIT  
LRRAVACLNHNTLTTFEDVKGHAHKAETALKMYMEFPQYNNNSVCSFLPEYTKR  
QTRDVTVALTTRHPWLSHTGDKPRDFTFKHFTFVMDIILDMSHNTLWYLCGS  
AFQMKGPFVPAVYKKWSAKGIQVGTWTVTFPEKSYSHSLSSYITDSEVDECEPH  
F"  
28025..29390  
/note="presence of CpG island"  
32335..60015  
/gene="327024.2"  
join(32335..32477,34498..34638,39032..39160,42555..44202,  
47273..47403,48499..48676,49233..49372,49490..51510,  
51409..51510,51706..51822,52429..52502,53024..53086,  
54567..54626,57808..60015)  
/gene="327024.2"  
join(34498..34638,39032..39160,42555..44202,47273..47403,  
48499..48676,49233..49372,49490..49606,51409..51510,  
51706..51822,52429..52502,53024..53086,54567..54626,  
57808..57883)  
/gene="327024.2"  
/codon\_start=1  
/product="KIAA0420"  
/protein\_id="AAC05804.1"  
/db\_xref="GI:2833634"  
/translation="MEYEFKCEKSLARTIOEASLTSESPFPAOSESTSLRFRGVAL  
SPLNIEKKEWQOKALDVEARKOVNKKALLTRVQSLIDNVQRPAPNSDFQ  
WEWETVNSSEVRLNVPATPNSPFSHTESHSPAKLDKTAGLIPEDNEQCKTDGID  
LARDSEGFNSPKCSDSNISHVENEAPEPKTSATPOTLISDPPSVNEQODPLILAE  
VTDDPYVMSLQNLKKSKKEYTEREOSRRSLRGSMNIVNESHLDKEDHDAVEVADCVES  
KQGLTGKRCVSLPDKRSKSNVLTQASLTQASMSMPPLASFKVDIPIRGHPV  
LESNSDFKVIPTVTENNVTKSLTGSFAPKSPSEPSNPMHRRSRSTTSACHILIN  
PINACEISPPGKQAMDLIQDPTENTNVEIEMPKLPTDIAGVCSKVYVGKATSEVY  
EDVLGKSNVCCSGNHLNKKYTHGLVTEGQLTSDERCAHNTSCAMPKLIHEVY  
ASSOCIAPNFGVSLKPSAMLEKNCSLQTELNKSYDVNPPLMONTQOMDT  
PMVSCNDEPLDPSFEKVRRLDIDIGLOKNCPIVITSGITEORHLPKRPYG  
SGVNRKRMKLTGTSKSEBELSKMLAFEMRRRLERQAOQSLLIABERERQ  
KEIEQKMLKEKKAHTAEASLDINNVALEWRKISDSLSLTMSQASLSITNSN  
SSGTSVAMQSVFVSNAPVSTSGSLTGLSVTRPGRKTRWSOVFSLEIQK  
FNKITAVAKGFLYRRLMQTDLKQLQRTVQDMETFRSFOSEAPLKGIVSADASIQ  
ERYLAQALRYLGHIDFFVMDAERMSILHHDREVREKELRQMDKSPRALSA  
TQKSLDKKKYKKAEMGMPIKPLVQNPSBETVLPQNGQNNPVHRLSLRQSSICK  
NPKKAACCDNLRHSLG"

ORIGIN  
Query Match 2.3%; Score 24; DB 9; Length 164564;  
Best Local Similarity 100.0%; Pred. No. 0.084;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 118 TGAGCCACACAGCTAGAAAGCAG 141  
123634 TGAGCCACACAGCTAGAAAGCAG 123657

RESULT 7  
AC137767  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC137767 86027 bp DNA linear PRI 21-JUN-2003  
Homo sapiens 12 BAC Pp13-942N8 (Roswell Park Cancer Institute Human  
BAC library) complete sequence.  
AC137767  
AC137767.16 GI:32140199  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;







```
repeat_region 5413..5720 /rpt_family="AluX"
repeat_region 5757..5857 /rpt_family="L1PA15-16"
repeat_region 5865..6155 /rpt_family="AluY"
repeat_region 6200..6490 /rpt_family="AluX"
repeat_region 6492..6635 /rpt_family="L1MD3"
repeat_region 6636..6942 /rpt_family="AluSg"
repeat_region 6943..6957 /rpt_family="L1MD3"
repeat_region 6960..7140 /rpt_family="AluY"
repeat_region 7173..7475 /rpt_family="AluSg"
repeat_region complement(7482..7568) /rpt_family="FLAM_A"
repeat_region 7569..7676 /rpt_family="L2"
repeat_region complement(8998..9293) /rpt_family="AluX"
repeat_region 9322..9630 /rpt_family="AluSg"
repeat_region 9752..9773 /rpt_family="AluSg"
repeat_region /rpt_family="AT-rich"
repeat_region complement(10253..10437) /rpt_family="MIR"
repeat_region 11149..11425 /rpt_family="AluYb"
repeat_region 11432..11734 /rpt_family="AluSg"
repeat_region 11759..11779 /rpt_family="AT-rich"
repeat_region 12892..13199 /rpt_family="AluX"
repeat_region 13236..13505 /rpt_family="L1MB7"
repeat_region 13506..13754 /rpt_family="AluX"
repeat_region 13755..14089 /rpt_family="L1MB7"
repeat_region 14118..14417
```

Query Match 2.1%; Score 22; DB 9; Length 86027;

Best Local Similarity 100.0%; Pred. No. 1.3;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 114 TGTCTGAGGCCACACAGCTAGA 135  
Db 2755 TGTCTGAGGCCACACAGCTAGA 2776

RESULT 8  
AL831755/c 91841 bp DNA linear PRI 26-SEP-2002  
LOCUS Human DNA sequence from clone RPI3-279N23 on chromosome 1, complete  
DEFINITION sequence.  
ACCESSION AL831755  
VERSION AL831755.9 GI:23337613  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 91841)  
REFERENCE  
AUTHORS Harrison, E.  
TITLE Direct Submission  
JOURNAL Submitted (26-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

## COMMENT

On Sep 27, 2002 this sequence version replaced gi:22798452.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPeP; Information on the WormPeP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RPI3-279N23 is from the library RPI3-13.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6.

## FEATURES

## Source

1..91841  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RPI3-279N23"  
/clone\_1lb="RPI3-13.2"

## ORIGIN

Query Match 2.1%; Score 22; DB 9; Length 91841;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 113 CTGCTGAGGCCACACAGCTAG 134  
Db 83423 CTGCTGAGGCCACACAGCTAG 83402

RESULT 9  
AL356748 97559 bp DNA linear PRI 04-DEC-2000  
LOCUS Human DNA sequence from clone RP4-686J16 on chromosome 1, complete  
DEFINITION sequence.  
ACCESSION AL356748  
VERSION AL356748.20 GI:11558567  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 97559)  
REFERENCE  
AUTHORS Laird, G.  
TITLE Direct Submission  
JOURNAL Submitted (03-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
On Dec 5, 2000 this sequence version replaced gi:11193283.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the correlation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em1, EMBL; SW, SWISSPROT; Tr, TREMBL; W, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping group. Further information can be found at <http://www.sanger.ac.uk/MSP/Chr1>

RP4-686316 is from the library RPCI-4 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP4-686316. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP4-686316 is at 1 in this sequence. The true left end of clone RP4-655N12 is at 97460 in this sequence. The true right end of clone RP4-78N12 is at 1573 in this sequence.

## FEATURES

**Source**

```

repeat_region
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-686J16"
/clone_11b="RPCI-4"
378. 514
/note="MIR repeat: matches 81. 226 of consensus"
927. 1084
/note="MIR repeat: matches 100. 258 of consensus"
1112. 1353
/note="AluX repeat: matches 1. 256 of consensus"
1512. 1812
/note="AluX repeat: matches 1. 300 of consensus"
Complement(1745. 2002)
/misc_feature
/note="match: GSS: Em:AQ321132"
1820. 1927
/note="AluYb repeat: matches 191. 298 of consensus"
2008. 2639
/note="match: GSS: Em:AQ474986"
2014. 2416
/note="match: GSS: Em:AQ706294"
2051. 2205
/note="MIR repeat: matches 78. 252 of consensus"
2328. 2414
/note="MLTH repeat: matches 51. 130 of consensus"
2477. 2555
/note="MLTH repeat: matches 202. 288 of consensus"
2676. 2813
/note="MLTF repeat: matches 401. 534 of consensus"
Complement(2908. 3284)
/misc_feature
/note="match: STS: Em:G18185"
3270. 3365
/note="L2 repeat: matches 2075. 2178 of consensus"
4118. 4212
/note="L1P3 repeat: matches 5725. 5822 of consensus"
7092. 7341
/note="match: STS: Em:G15587"
Complement(7568. 8088)
/misc_feature
/note="match: GSS: Em:AQ707880"
Complement(7909. 8281)
/misc_feature
/note="match: GSS: Em:AQ035467"
8073. 8228
repeat_region

```

repeat_region	/note="MIR repeat: matches 106. .245 of consensus"
repeat_region	8286. .8342 /note="L2 repeat: matches 2691. .2749 of consensus"
repeat_region	8660. .8744 /note="L2 repeat: matches 2217. .2298 of consensus"
repeat_region	8745. .9049 /note="AluY repeat: matches 1. .305 of consensus"
misc_feature	9050. .9230 /note="L2 repeat: matches 2023. .2217 of consensus"
misc_feature	10540. /note="match: GSS: Em:AQ316052"
repeat_region	10633. /note="match: GSS: Em:A2121092"
repeat_region	11304. .11419 /note="LTRB repeat: matches 445. .560 of consensus"
repeat_region	12397. .12857 /note="LTRG repeat: matches 31. .512 of consensus"
misc_feature	13076. /note="match: GSS: Em:AQ743494"
repeat_region	13492. .13591 /note="L2 repeat: matches 2388. .2487 of consensus"
repeat_region	13685. .13725 /note="L2 repeat: matches 2709. .2749 of consensus"
repeat_region	14198. .14273 /note="L2 repeat: matches 2673. .2749 of consensus"
repeat_region	15025. .15148 /note="L2 repeat: matches 2584. .2708 of consensus"
repeat_region	16477. .16612 /note="MIR repeat: matches 85. .228 of consensus"
misc_feature	16832. /note="match: GSS: Em:AQ700892"
repeat_region	17109. .17167 /note="MER45B repeat: matches 437. .492 of consensus"
misc_feature	17344. .17892 /note="match: GSS: Em:B50102"
repeat_region	17496. .17773 /note="AluSq repeat: matches 37. .311 of consensus"
repeat_region	18470. .18669 /note="L1MC/D repeat: matches 5423. .5612 of consensus"
repeat_region	18751. .18902 /note="L1MC5 repeat: matches 7758. .7913 of consensus"
repeat_region	18928. .19233 /note="AluX repeat: matches 5. .312 of consensus"
repeat_region	19235. .19296 /note="31 copies 2 mex aa 74% conserved"
repeat_region	19553. .19654 /note="AluJc/FRM repeat: matches 203. .304 of consensus"
repeat_region	19652. .19772 /note="AluJ/LAM repeat: matches 5. .85 of consensus"
repeat_region	20881. .20991 /note="MIR repeat: matches 16. .142 of consensus"
repeat_region	21460. .21617 /note="MIR repeat: matches 20. .184 of consensus"
repeat_region	22616. .23224 /note="L1ME1 repeat: matches 5528. .6159 of consensus"
repeat_region	23288. .23435 /note="MIR repeat: matches 80. .233 of consensus"
repeat_region	25186. .25354 /note="MIR repeat: matches 77. .262 of consensus"
repeat_region	25530. .25827 /note="AluX repeat: matches 1. .298 of consensus"
misc_feature	26592. /note="match: GSS: Em:AQ532789"
misc_feature	26598. .27124 /note="match: GSS: Em:AQ619294"
misc_feature	26720. .27017 /note="match: GSS: Em:AQ100816"
misc_feature	26833. .27311 /note="match: GSS: Em:AQ837606"
repeat_region	26971. .27041 /note="MTRUJ repeat: matches 110. .180 of consensus"
misc_feature	27111. .27147 /note="Single clone region. Assembly confirmed by

restriction digest data."

misc\_feature 27131..27224  
/note="match: GSS: Em:AQ136115"  
27133..27439  
/note="match: GSS: Em:AQ55613"  
27139..27631  
/note="match: GSS: Em:AQ798882"  
27202..27252  
/note="3 copies 17 mer 86% conserved"  
misc\_feature 27218..27578  
/note="match: GSS: Em:AQ438628"  
27236..27568  
/note="match: GSS: Em:AQ136115"  
29394..29510  
/note="L2 repeat: matches 2587..2699 of consensus"  
29562..29638  
/note="L2 repeat: matches 2587..2664 of consensus"  
32900..32937  
/note="L2 repeat: matches 2660..2705 of consensus"  
33091..33356  
/note="match: GSS: Em:AQ16053"  
33999..34281  
/note="L1 repeat: matches 17..311 of consensus"  
34351..34640  
/note="L1 repeat: matches 1..287 of consensus"  
35737..36047  
/note="L1 repeat: matches 1..313 of consensus"  
37307..37627  
/note="L1 repeat: matches 1..293 of consensus"  
38057..38173  
/note="L1 repeat: matches 1..118 of consensus"  
38174..38190  
/note="L1 repeat: matches 1555..1871 of consensus"  
38191..38449  
/note="L1 repeat: matches 1..302 of consensus"

Query Match 2.1%; Score 22; DB 9; Length 97559;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

113 CTGCTGAGGCCACAGCTAG 134  
Db 20910 CTGCTGAGGCCACAGCTAG 20931

RESULT 10  
AL157934/c 106497 bp DNA linear PRI 06-JAN-2002  
LOCUS Human DNA sequence from clone Rpl1-449M9 on chromosome Xq13.1-13.3,  
DEFINITION complete sequence.  
ACCESSION AL157934  
VERSION AL157934.17 GI:18077665  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 106497)  
AUTHORS Wray, P.  
TITLE Direct Submission  
JOURNAL Submitted (05-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk; Clone requests: clonerequests@sanger.ac.uk  
On Jan 6, 2002 this sequence version replaced gi:17381304.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em, EMBL; Sw,  
SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormPep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome X, constructed by the Sanger Centre Chromosome X Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/ChrX  
Rpl1-449M9 is from the library RPl1-11.2 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6  
IMPORTANT: This sequence is not the entire insert of clone  
Rpl1-449M9 it may be shorter because we sequence overlapping  
sections only once, except for a short overlap.  
The true left end of clone CTD-2530H13 is at 104498 in this  
sequence. The true right end of clone XX-BMXD718 is at 2000 in this  
sequence.

FEATURES  
source  
1..106497  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="q13.1-13.3"  
/clone="Rpl1-449M9"  
/clone\_1id="RPl1-11.2"  
10905..11106  
/note="Sequence from clone PCR only."

Query Match 2.1%; Score 22; DB 9; Length 106497;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

931 ACAGGGGCTCTCAGAGAGCCT 952  
Db 19854 ACAGGGGCTCTCAGAGAGCCT 19833

RESULT 11  
CO869788/c 112453 bp DNA linear PAT 13-SEP-2004  
LOCUS Sequence 209 from Patent WO2004074320.  
DEFINITION CO869788  
ACCESSION CO869788  
VERSION CO869788.1 GI:51999587  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Morris, D.W., Morris, D.W. and Malandro, M.S.  
TITLE Novel therapeutic targets in cancer  
JOURNAL Patent: WO 2004074320-A-209 02-SEP-2004;  
Sagres Discovery, Inc. (US)  
FEATURES  
source  
1..112453  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

Query Match 2.1%; Score 22; DB 6; Length 112453;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

113 CTGCTGAGGCCACAGCTAG 134

Db 33503 CTGTGAGCCACACAGCTAG 33482

RESULT 12 AC137629 125836 bp DNA linear HTG 05-APR-2003  
LOCUS Homo sapiens clone RP11-525N13, WORKING DRAFT SEQUENCE.

AC137629 AC137629.16 GI:29501835

VERSION HTG, HTGS, PHASE2, HTGS\_DRAFT, HTGS\_FULLTOP, HTGS\_ACTIVEFIN.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 125836)

REFERENCE  
AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Albrooks, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banks, T., Barbata, J., Benton, J., Bimane, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowle, S., Bivela, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.E., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H., Douhwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, J., Koryah, J., Kovar, C., Kratovic, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichteberg, O., Lieu, C., Liu, J., Liu, W., Loulesged, H., Locado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabadi, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenwo, S., Ogih, M., Okumura, G., Oreguine, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojao, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Slabson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sultan, A., Svatek, A., Tabor, P., Tameria, A., Tameria, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watling, S., Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

COMMENT

Submitted (05-APR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Apr 3, 2003 this sequence version replaced gi:2942847.  
----- Genome Center  
Center: Baylor College of Medicine

Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: hgsc-help@bcm.tmc.edu

Project Information  
Center project name: HERN  
Center clone name: RP11-525N13

Summary Statistics  
Sequencing vector: Plasmid;  
Sequencing vector: M13;

Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 125202 bases at least Q40  
Consensus quality: 125396 bases at least Q30  
Consensus quality: 125495 bases at least Q20

Estimated insert size: 136468; sum-of-configs estimation  
Quality coverage: 20x in Q20 bases; sum-of-configs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
\* NOTE: This is a 'near complete' sequence.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 125836: contigs of 125836 bp in length.  
Location/Qualifiers  
1..125836

FEATURES  
source  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-525N13"

misc\_feature  
41073..71172  
/note="remaining sized gap in near\_complete"

misc\_feature  
90658..90857  
/note="remaining sized gap in near\_complete"

ORIGIN

Query Match 2.1%; Score 22; DB 2; Length 125836;  
Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 TGTCTGAGCCACACAGCTAGA 135

Db 56334 TGTCTGAGCCACACAGCTAGA 56313

RESULT 13

LOCUS AC002419/c

DEFINITION Homo sapiens chromosome X clone bKMD40, \*\*\* SEQUENCING IN PROGRESS  
\*\*\* 2 unordered pieces.

ACCESSION AC002419  
VERSION AC002419.2 GI:21405639

KEYWORDS HTG, HTGS, PHASE1.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
1 (bases 1 to 128440)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Chen, B., Brownstein, B.H., States, D.J., Schlessinger, D. and  
Mazarella, R.

Direct Submission  
Unpublished (1997)

REFERENCE 2 (bases 1 to 128440)  
Brownstein, B.H., States, D.J. and Mazarella, R.  
Direct Submission  
Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,

## COMMENT

Washington University School of Medicine, 4566 Scott Avenue, St. Louis, MO 63110, USA  
 On Jun 13, 2002 this sequence version replaced gi:2333250.  
 Current status of this project is available at:  
 'http://genome.wustl.edu/cgm/seq\_projects.html'  
 Submitted by:  
 Ellison Chen,  
 Advanced Center for Genetic Technology,  
 Applied Biosystems Division of Perlin Elmer Corp.,  
 850 Lincoln Center Drive,  
 Foster City, CA 94404 USA  
 e-mail: ellison@genseq.aplbdio.com

and

Buddy Brownstein,  
 Center for Genetics in Medicine,  
 Washington University School of Medicine, Box 8232  
 4566 Scott Avenue,  
 St. Louis, MO 63110, USA  
 e-mail: buddy@genetics.wustl.edu

and

David J. States,  
 Institute for Biomedical Computing  
 Washington University in St. Louis  
 700 South Euclid Ave.  
 St. Louis, MO 63108 USA  
 e-mail: states@ibc.wustl.edu.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 58852: contig of 58852 bp in length  
 \* 58853 58952: gap of 100 bp  
 \* 58953 128440: contig of 69488 bp in length.

## FEATURES

## SOURCE

1. 128440  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /clone="bmx40"

## ORIGIN

Query Match 2.1%; Score 22; DB 2; Length 128440;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 931 ACAGGGCTCTCAGAGAGGCT 952

Db 103762 ACAGGGCTCTCAGAGAGGCT 103741

## RESULT 14

AL844513 131184 bp DNA linear ROD 18-NOV-2002  
 LOCUS Mouse DNA sequence from clone RP23-132A3 on chromosome 2, complete  
 DEFINITION sequence.

ACCESSION AL844513  
 VERSION AL844513.8 GI:33438736

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 131184)  
 Mammalia; Eutheria; Chordata; Sciuromorphi; Muridae; Murinae; Mus.  
 AUTHORS Johnson, C.

## TITLE

## JOURNAL

## COMMENT

Direct Submission  
 Submitted (19-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Aug 5, 2003 this sequence version replaced gi:24816942.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 EM, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wg, WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-132A3 is  
 from the RPI-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see http://www.chori.org/bacpac/home.htm  
 VECTOR: pBac3.6.

## FEATURES

## SOURCE

1. 131184  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP23-132A3"  
 /clone\_lib="RPI-23"

## ORIGIN

Query Match 2.1%; Score 22; DB 10; Length 131184;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 AGTCCATTTCAGATGAGGA 86

Db 2737 AGTCCATTTCAGATGAGGA 2758

## RESULT 15

AC104081 142118 bp DNA linear PRI 01-MAR-2002  
 LOCUS Homo sapiens BAC clone RP11-559A1 from 2, complete sequence.  
 DEFINITION AC104081 AC068259

ACCESSION AC104081.2 GI:117977480

VERSION HTG.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 142118)  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 Sultson, J.E. and Waterston, R.  
 Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792

PUBMED 9847074  
 2. (bases 1 to 142118)  
 REFERENCE  
 AUTHORS Iwak, A. and Meyer, R.  
 TITLE The sequence of Homo sapiens BAC clone Rpl1-559A1  
 JOURNAL Unpublished (2001)  
 REFERENCE  
 3. (bases 1 to 142118)  
 AUTHORS  
 TITLE Direct Submission  
 JOURNAL Submitted (03-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE  
 4. (bases 1 to 142118)  
 AUTHORS  
 TITLE Direct Submission  
 JOURNAL Submitted (22-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE  
 5. (bases 1 to 142118)  
 AUTHORS  
 TITLE Direct Submission  
 JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE  
 6. (bases 1 to 142118)  
 AUTHORS  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Dec 22, 2001 this sequence version replaced gi:17227302.  
 -----  
 Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
 -----  
 Summary Statistics  
 Center project name: H\_NH0559A01  
 Drafting Center: WIBR  
 -----  
 NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.  
 This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., paired quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.  
 MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>  
 SOURCE INFORMATION:  
 The RPl1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, B., Tatero, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBACe3.6  
 NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is Rpl1-50B16; the clone sequenced to the right is Rpl1-554H10, 2000 bp overlap. Actual end of this clone is at base position 79619 of Rpl1-554H10.

Data from AC012671 was used to finish this clone, AC068259. Polymorphisms have been identified between AC012671 and AC068259.  
 The sequence of AC068259 has been incorporated into AC104081.  
 FEATURES  
 source

Location/Qualifiers  
 1..142118  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="Rpl1-559A1"  
 /clone\_1fb="RPl1-11"  
 1..964  
 /rpt\_family="L1"  
 repeat\_region  
 971..1204  
 /rpt\_family="L1"  
 repeat\_region  
 976..1036  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 1073..1114  
 /rpt\_family="TATTA)n"  
 repeat\_region  
 1243..1518  
 /rpt\_family="MER1\_type"  
 repeat\_region  
 1271..1338  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 1553..1580  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 1814..1891  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 2254..2741  
 /rpt\_family="L2"  
 repeat\_region  
 3256..3311  
 /rpt\_family="MALR"  
 repeat\_region  
 4153..4187  
 /rpt\_family="MIR"  
 repeat\_region  
 4234..4317  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 4750..4793  
 /rpt\_family="ERVL"  
 repeat\_region  
 5564..5620  
 /rpt\_family="L2"  
 repeat\_region  
 6370..6404  
 /rpt\_family="MIR"  
 repeat\_region  
 6550..6681  
 /rpt\_family="(CA)n"  
 repeat\_region  
 6704..7158  
 /rpt\_family="MIR"  
 repeat\_region  
 7009..7031  
 /rpt\_family="ERVL"  
 repeat\_region  
 7189..7354  
 /rpt\_family="(T)n"  
 repeat\_region  
 7650..7692  
 /rpt\_family="(ATG)n"  
 repeat\_region  
 7929..7996  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 7984..8103  
 /rpt\_family="MIR"  
 repeat\_region  
 8425..8461  
 /rpt\_family="MIR"  
 repeat\_region  
 8543..8588  
 /rpt\_family="MIR"  
 repeat\_region  
 8637..9095  
 /rpt\_family="MALR"  
 repeat\_region  
 9178..9260  
 /rpt\_family="MIR"  
 repeat\_region  
 9684..9808  
 /rpt\_family="MIR"  
 repeat\_region  
 10301..10341  
 /rpt\_family="MIR"  
 repeat\_region  
 10978..11002  
 /rpt\_family="(CA)n"  
 repeat\_region  
 11305..11402  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 /rpt\_family="MIR"

```

repeat_region      12182. .12244
/rpc_family="Alu"
repeat_region      12245. .12587
/rpc_family="MALR"
repeat_region      14154. .14485
/rpc_family="MALR"
repeat_region      14486. .14725
/rpc_family="Alu"
repeat_region      14704. .14749
/rpc_family="(TAAA)n"
repeat_region      15138. .15429
/rpc_family="Alu"
repeat_region      15811. .16068
/rpc_family="MER1_type"
repeat_region      16065. .16458
/rpc_family="MALR"
repeat_region      16459. .16530
/rpc_family="MER1_type"
repeat_region      17696. .18200
/rpc_family="ERV1"
repeat_region      18458. .18597
/rpc_family="MALR"
repeat_region      18598. .19184
/rpc_family="MALR"
repeat_region      18943. .18997
/rpc_family="GA-rich"
repeat_region      19185. .19515
/rpc_family="MALR"
repeat_region      19593. .20149
/rpc_family="MALR"
repeat_region      20252. .20416
/rpc_family="L2"
repeat_region      20842. .20881
/rpc_family="(TTTC)n"
repeat_region      20853. .21166
/rpc_family="Alu"
repeat_region      21410. .21551
/rpc_family="MIR"
repeat_region      22386. .22535
/rpc_family="CT-rich"

```

Query Match 2.1%; Score 22; DB 9; Length 142118;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 AGTCCATTTCAGATGAGA 86  
 ||||||||||||||||  
 Db 24601 AGTCCATTTCAGATGAGA 24622

Search completed: June 29, 2005, 19:59:07  
 Job time : 4895 secs

**THIS PAGE BLANK (USPTO)**